IDENTIFICATION, ANALYSIS AND CONTROL OF DYNAMICAL SYSTEMS

PART 2: SYSTEMS IDENTIFICATION

Alberto Bemporad

Academic year 2016-2017
System identification: introduction
The design of a controller/observer requires a mathematical model describing the behaviour of the plant.
The design of a controller/observer requires a mathematical model describing the behaviour of the plant.

A model describes how the signals of the system are related to each other.
The design of a controller/observer requires a mathematical model describing the behaviour of the plant.

A model describes how the signals of the system are related to each other.

A model and a system are two different objects.
Building mathematical models

- The design of a controller/observer requires a **mathematical model** describing the behaviour of the plant.
- A **model describes** how the signals of the system are related to each other.
- A model and a system are two different objects.
- Different kinds of models:
  - **Mental or intuitive** models. For example:
    - when driving a car, pushing the break decreases the speed.
  - **Graphical** models. For example:
    - Bode diagram or step response of an LTI system;
    - current-voltage characteristic of a diode.
  - **Mathematical** models, described by equations.
The design of a controller/observer requires a **mathematical model** describing the behaviour of the plant.

A **model describes** how the signals of the system are related to each other.

A model and a system are two different objects.

Different kinds of models:
- **mental or intuitive** models. For example: when driving a car, pushing the break decreases the speed.
- **graphical** models. For example: Bode diagram or step response of an LTI system; current-voltage characteristic of a diode.
- **mathematical** models, described by equations.

We will focus on mathematical models of dynamical systems, described, in general, by **differential or difference equations**.
The design of a controller/observer requires a **mathematical model**
describing the behaviour of the plant.

A **model describes** how the signals of the system are related to each other.

A model and a system are two different objects.

Different kinds of models:

- **mental or intuitive** models. For example:
  - when driving a car, pushing the break decreases the speed.
- **graphical** models. For example:
  - Bode diagram or step response of an LTI system;
  - current-voltage characteristic of a diode.
- **mathematical** models, described by equations.

We will focus on mathematical models of dynamical systems, described, in
general, by **differential or difference equations**.

Mathematical models can be derived from:

- **first principle laws of physics, chemistry, biology, etc.**
  (physical modeling approach)
- observed data generated by the system
  (system identification approach)
The system identification procedure involves three basic entities:

1. **Data**, which can be either recorded from specifically designed experiments or from normal operations of the system.

2. **Set of candidate models**, obtained by specifying within which set of models we are going to look for a suitable one. Different kinds of models may be specified (e.g., linear vs nonlinear; continuous time vs discrete time; deterministic vs stochastic, etc.). Two types of model sets:
   - **gray boxes**. A model with some unknown parameters is derived from physical laws. The parameters are then estimated from data.
   - **black boxes**. A model structure is chosen (e.g., linear models). The parameters of the model do not reflect any physical consideration.

3. **Rule to assess candidate models** using data. This is the identification method, used to determinate the “best” model in the set, guided by data.
Test whether the estimated model is an “appropriate” representation of the system. Assess how the model relates to:

- **prior knowledge**. Does the model adequately describes prior known physical behaviour of the system?
- **experimental data** (not used for training). Compare the simulated outputs of the model with the observed outputs.
System identification loop

1. Design of experiment
2. Perform experiment
3. Collect data
4. Choose model structure
5. Estimate the model
6. Validate the model

If the model is NOT accepted, go back to choose model structure. If the model is accepted, use it.

Input: Prior knowledge

Output: New data


LTI systems
Given a discrete-time signal $u(k)$, $k = 0, 1, \ldots$, we define the (unilateral) $z$-transform of $u$ as

$$
\mathcal{Z}\{u(k)\} = U(z) = \sum_{k=0}^{\infty} u(k)z^{-k}
$$

- $\mathcal{Z}\{u(k - d)\} = \mathcal{Z}\{q^{-d}u(k)\} = z^{-d}U(z), \quad d \in \mathbb{Z}$
- $\mathcal{Z} \{g(k) \ast u(k)\} = \mathcal{Z} \left\{ \sum_{\ell=0}^{\infty} g(\ell)u(k - \ell) \right\} = \mathcal{Z} \{G(q)u(k)\} = G(z)U(z)$

![Diagram of input/output representation](image URL)
Given a discrete-time signal $u(k)$, $k = 0, 1, \ldots$, we define the (unilateral) $z$-transform of $u$ as

$$
\mathcal{Z}\{u(k)\} = U(z) = \sum_{k=0}^{\infty} u(k)z^{-k}
$$

- $\mathcal{Z}\{u(k - d)\} = \mathcal{Z}\{q^{-d}u(k)\} = z^{-d}U(z)$, $d \in \mathbb{Z}$
- $\mathcal{Z}\{g(k) \ast u(k)\} = \mathcal{Z}\left\{\sum_{\ell=0}^{\infty} g(\ell)u(k - \ell)\right\} = \mathcal{Z}\{G(q)u(k)\} = G(z)U(z)$

Analogy between the time-domain operator $G(q)$ and the DT transfer function $G(z)$
Given a discrete-time signal \( u(k), k = 0, 1, \ldots \), we define the (unilateral) z-transform of \( u \) as

\[
\mathcal{Z}\{u(k)\} = U(z) = \sum_{k=0}^{\infty} u(k)z^{-k}
\]

- \( \mathcal{Z}\{u(k - d)\} = \mathcal{Z}\{q^{-d}u(k)\} = z^{-d}U(z), \quad d \in \mathbb{Z} \)
- \( \mathcal{Z}\{g(k) \ast u(k)\} = \mathcal{Z}\left\{ \sum_{\ell=0}^{\infty} g(\ell)u(k - \ell) \right\} = \mathcal{Z}\{G(q)u(k)\} = G(z)U(z) \)

Analogy between the time-domain operator \( G(q) \) and the DT transfer function \( G(z) \)

Thanks to this analogy, we can treat \( G(q) \) as polynomials in \( q \). Product and ratio between \( G_1(q) \) and \( G_2(q) \) have a meaning!

Example: \( y(k) = \frac{b_1q^{-1}}{1+a_1q^{-1}} u(k) \rightarrow (1 + a_1q^{-1})y(k) = b_1q^{-1}u(k) \)
Linear regression representation

- Linear regression representation of the system:

\[
    y(k) = \varphi^\top(k)\theta
\]

\(\theta\): parameter vector, \(\varphi(k)\): regressor vector, typically containing past values of inputs and outputs.

\[
    \varphi(k) = [-y(k-1) \ldots -y(k-n_a) \ u(k) \ldots u(k-n_b)]^\top
\]

\[
    \theta = [a_1 \ldots a_{n_a} \ b_0 \ldots b_{n_b}]^\top
\]

Writing out the product gives:

\[
    y(k) = G(q)u(k), \quad G(q) = \frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}}
\]

Non-linear systems can be easily represented in a linear regression form. Just include nonlinear terms (e.g., \(y^2(k-1); u(k)y(k-1)\)) in the regressor!
Least-squares estimation
Consider a model in the linear regression form: $\mathcal{M} : \hat{y}(k, \theta) = \phi^\top(k)\theta$
Consider a model in the linear regression form: \( M : \hat{y}(k, \theta) = \varphi^\top(k)\theta \)

Define the residuals as
\[
\varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta) = y(k) - \varphi^\top(k)\theta
\]

\( \varepsilon(k, \theta) \) represents the error between output observations and model outputs \( \hat{y}(k, \theta) \)
Consider a model in the linear regression form: $\mathcal{M}: \hat{y}(k, \theta) = \varphi^T(k)\theta$

Define the residuals as $\varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta) = y(k) - \varphi^T(k)\theta$

$\varepsilon(k, \theta)$ represents the error between output observations and model outputs $\hat{y}(k, \theta)$

Least-squares (LS) estimate:

$$\hat{\theta}_{LS} = \arg\min_{\theta} \sum_{k=1}^{N} \varepsilon^2(k, \theta) = \arg\min_{\theta} \sum_{k=1}^{N} \left( y(k) - \varphi^T(k)\theta \right)^2$$
Consider a model in the linear regression form: $\mathcal{M} : \hat{y}(k, \theta) = \varphi^\top(k)\theta$

Define the residuals as $\varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta) = y(k) - \varphi^\top(k)\theta$

$\varepsilon(k, \theta)$ represents the error between output observations and model outputs $\hat{y}(k, \theta)$

Least-squares (LS) estimate:

$$\hat{\theta}_{LS} = \arg\min_{\theta} \sum_{k=1}^{N} \varepsilon^2(k, \theta) = \arg\min_{\theta} \sum_{k=1}^{N} \left( y(k) - \varphi^\top(k)\theta \right)^2 = \arg\min_{\theta} \| Y - \Phi \theta \|^2$$

$$Y = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad \Phi = \begin{bmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{bmatrix}$$
Consider a model in the linear regression form: 
\[ M : \hat{y}(k, \theta) = \varphi^\top(k)\theta \]

Define the residuals as 
\[ \varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta) = y(k) - \varphi^\top(k)\theta \]

\( \varepsilon(k, \theta) \) represents the error between output observations and model outputs \( \hat{y}(k, \theta) \)

Least-squares (LS) estimate:
\[
\hat{\theta}_{\text{LS}} = \arg \min_{\theta} \sum_{k=1}^{N} \varepsilon^2(k, \theta) = \arg \min_{\theta} \sum_{k=1}^{N} \left( y(k) - \varphi^\top(k)\theta \right)^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2
\]

\[
Y = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad \Phi = \begin{bmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{bmatrix}
\]

Solution of the QP problem:
\[
\hat{\theta}_{\text{LS}} : \frac{\partial \| Y - \Phi \theta \|^2}{\partial \theta} = 0
\]
Linear least-squares

- Consider a model in the linear regression form: \( M : \hat{y}(k, \theta) = \varphi^\top(k)\theta \)
- Define the residuals as \( \varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta) = y(k) - \varphi^\top(k)\theta \)

\( \varepsilon(k, \theta) \) represents the error between output observations and model outputs \( \hat{y}(k, \theta) \)

- Least-squares (LS) estimate:

\[
\hat{\theta}_{\text{LS}} = \arg \min_{\theta} \sum_{k=1}^{N} \varepsilon^2(k, \theta) = \arg \min_{\theta} \sum_{k=1}^{N} \left( y(k) - \varphi^\top(k)\theta \right)^2 = \arg \min_{\theta} \| Y - \Phi\theta \|^2
\]

\[
Y = \begin{bmatrix}
y(1) \\
\vdots \\
y(N)
\end{bmatrix}, \quad \Phi = \begin{bmatrix}
\varphi^\top(1) \\
\vdots \\
\varphi^\top(N)
\end{bmatrix}
\]

- Solution of the QP problem:

\[
\hat{\theta}_{\text{LS}} : \frac{\partial \| Y - \Phi\theta \|^2}{\partial \theta} = 0 \rightarrow \hat{\theta}_{\text{LS}} = \left( \Phi^\top\Phi \right)^{-1} \Phi^\top Y = \left( \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \sum_{k=1}^{N} \varphi(k)y(k)
\]
Consider a model in the linear regression form: \( M : \hat{y}(k, \theta) = \varphi^\top(k)\theta \)

Define the residuals as \( \varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta) = y(k) - \varphi^\top(k)\theta \)

\( \varepsilon(k, \theta) \) represents the error between output observations and model outputs \( \hat{y}(k, \theta) \)

Least-squares (LS) estimate:

\[
\hat{\theta}_{LS} = \arg \min_{\theta} \sum_{k=1}^{N} \varepsilon^2(k, \theta) = \arg \min_{\theta} \sum_{k=1}^{N} \left( y(k) - \varphi^\top(k)\theta \right)^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2
\]

\[
Y = \begin{bmatrix}
y(1) \\
\vdots \\
y(N)
\end{bmatrix}, \quad \Phi = \begin{bmatrix}
\varphi^\top(1) \\
\vdots \\
\varphi^\top(N)
\end{bmatrix}
\]

Solution of the QP problem:

\[
\hat{\theta}_{LS} : \frac{\partial \| Y - \Phi \theta \|^2}{\partial \theta} = 0 \rightarrow \hat{\theta}_{LS} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y = \left( \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \sum_{k=1}^{N} \varphi(k)y(k)
\]

Matlab: \( \hat{\theta}_{LS} = \Phi \backslash Y \)
\[ \hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \]

- \( \hat{\theta}_{LS} \) is the solution of the set of linear equations:
  \[ \Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y \]
Linear least-squares: Cholesky factorization

\[ \hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \]

- \( \hat{\theta}_{LS} \) is the solution of the set of linear equations:
  \[ \Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y \]

- Use Cholesky decomposition of \( \Phi^T \Phi \) to solve the above system of linear equations, i.e.,
  \[ \Phi^T \Phi = LL^T \quad L : \text{Lower Triangular Matrix} \]
Linear least-squares: Cholesky factorization

\[ \hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \]

- \( \hat{\theta}_{LS} \) is the solution of the set of linear equations:
  \[ \Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y \]

- Use Cholesky decomposition of \( \Phi^T \Phi \) to solve the above system of linear equations, i.e.,
  \[ \Phi^T \Phi = LL^T \quad L : \text{Lower Triangular Matrix} \]

Matlab: \( L = \text{chol}(\Phi^T \Phi,'lower') \)
Linear least-squares: Cholesky factorization

\[ \hat{\theta}_{\text{LS}} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- \( \hat{\theta}_{\text{LS}} \) is the solution of the set of linear equations:
  \[ \Phi^\top \Phi \hat{\theta}_{\text{LS}} = \Phi^\top Y \]

- Use Cholesky decomposition of \( \Phi^\top \Phi \) to solve the above system of linear equations, i.e.,
  \[ \Phi^\top \Phi = LL^\top \quad L : \text{Lower Triangular Matrix} \]

  Matlab: \( L = \text{chol}(\Phi^\top \Phi, 'lower') \)

- Least-squares (LS) estimate:
  \[ LL^\top \hat{\theta}_{\text{LS}} = \Phi^\top Y \]
Linear least-squares: Cholesky factorization

\[ \hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \]

- \( \hat{\theta}_{LS} \) is the solution of the set of linear equations:
  \[ \Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y \]

- Use Cholesky decomposition of \( \Phi^T \Phi \) to solve the above system of linear equations, i.e.,
  \[ \Phi^T \Phi = LL^T \]
  \( L \): Lower Triangular Matrix
  
  Matlab: \( L = \text{chol}(\Phi^T \Phi, \text{'lower'}) \)

- Least-squares (LS) estimate:
  \[
  LL^T \hat{\theta}_{LS} = \Phi^T Y \\
  \downarrow \\
  Lz = \Phi^T Y, \quad L^T \hat{\theta}_{LS} = z
  \]
Linear least-squares: Cholesky factorization

\[ \hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \]

- \( \hat{\theta}_{LS} \) is the solution of the set of linear equations:
  \[ \Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y \]

- Use Cholesky decomposition of \( \Phi^T \Phi \) to solve the above system of linear equations, i.e.,
  \[ \Phi^T \Phi = LL^T \]
  \( L \): Lower Triangular Matrix

  **Matlab:** \( L = \text{chol}(\Phi^T \Phi, 'lower') \)

- Least-squares (LS) estimate:
  \[ LL^T \hat{\theta}_{LS} = \Phi^T Y \]

  \[ \downarrow \]

  \[ Lz = \Phi^T Y, \quad L^T \hat{\theta}_{LS} = z \]

- Solve the linear system \( Lz = \Phi^T Y \) through **forward** substitution
\[ \hat{\theta}_{LS} = (\Phi^T \Phi)^{-1} \Phi^T Y \]

- \( \hat{\theta}_{LS} \) is the solution of the set of linear equations:
  \[ \Phi^T \Phi \hat{\theta}_{LS} = \Phi^T Y \]

- Use Cholesky decomposition of \( \Phi^T \Phi \) to solve the above system of linear equations, i.e.,
  \[ \Phi^T \Phi = LL^T \quad L : \text{Lower Triangular Matrix} \]

Matlab: \( L = \text{chol}(\Phi^T \Phi,'lower') \)

- Least-squares (LS) estimate:
  \[ LL^T \hat{\theta}_{LS} = \Phi^T Y \]
  \[ \downarrow \]
  \[ Lz = \Phi^T Y, \quad L^T \hat{\theta}_{LS} = z \]

- Solve the linear system \( Lz = \Phi^T Y \) through forward substitution
- Solve the linear system \( L^T \hat{\theta}_{LS} = z \) through backward substitution
$\hat{\theta}_{LS} = \left( \Phi^{\top} \Phi \right)^{-1} \Phi^{\top} Y$

- Compute a QR factorization of the (full-column rank) matrix $\Phi \in \mathbb{R}^{N,n}$, i.e.,

$\Phi = \begin{bmatrix} [Q_1]_{N,n} & [Q_2]_{N,(N-n)} \end{bmatrix} \begin{bmatrix} [R_1]_{n,n} \\ 0_{N-n,n} \end{bmatrix} \underbrace{Q}_{R}$

with $Q_1^{\top} Q_1 = I$, $R_1$ upper triangular, $r_{ii} > 0$ if $\Phi$ is full-column rank.
Linear least-squares: QR factorization

\[ \hat{\theta}_{LS} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- Compute a QR factorization of the (full-column rank) matrix \( \Phi \in \mathbb{R}^{N,n} \), i.e.,

\[
\Phi = \begin{bmatrix} [Q_1]_{N,n} & [Q_2]_{N,(N-n)} \end{bmatrix} \underbrace{\begin{bmatrix} [R_1]_{n,n} \\ 0_{N-n,n} \end{bmatrix}}_{R}
\]

Matlab: `qr(\Phi)`

with \( Q_1^\top Q_1 = I \), \( R_1 \) upper triangular, \( r_{ii} > 0 \) if \( \Phi \) is full-column rank.
Linear least-squares: QR factorization

\[ \hat{\theta}_{LS} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- Compute a QR factorization of the (full-column rank) matrix \( \Phi \in \mathbb{R}^{N,n} \), i.e.,

\[
\Phi = \begin{bmatrix} [Q_1]_{N,n} & [Q_2]_{N,(N-n)} \end{bmatrix} \begin{bmatrix} [R_1]_{n,n} \\ 0_{N-n,n} \end{bmatrix}
\]

Matlab: `qr(\( \Phi \))`

- with \( Q_1^\top Q_1 = I \), \( R_1 \) upper triangular, \( r_{ii} > 0 \) if \( \Phi \) is full-column rank.

- Substitution:

\[
\hat{\theta}_{LS} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y = \left( (Q_1 R_1)^\top (Q_1 R_1) \right)^{-1} (R_1^\top Q_1^\top Y) = \left( R_1^\top Q_1^\top Q_1 R_1 \right)^{-1} R_1^\top Y
\]

\[
= \left( R_1^\top R_1 \right)^{-1} R_1^\top Q_1^\top Y = R_1^{-1} R_1^{-\top} R_1^\top Q_1^\top Y = R_1^{-1} Q_1^\top Y
\]
Linear least-squares: QR factorization

\[ \hat{\theta}_{LS} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- Compute a QR factorization of the (full-column rank) matrix \( \Phi \in \mathbb{R}^{N,n} \), i.e.,

\[ \Phi = \begin{bmatrix} [Q_1]_{N,n} & [Q_2]_{N,(N-n)} \end{bmatrix} \begin{bmatrix} [R_1]_{n,n} \\ 0_{N-n,n} \end{bmatrix} \]

with \( Q_1^\top Q_1 = I \), \( R_1 \) upper triangular, \( r_{ii} > 0 \) if \( \Phi \) is full-column rank.

- Substitution:

\[ \hat{\theta}_{LS} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y = \left( (Q_1 R_1)^\top (Q_1 R_1) \right)^{-1} (R_1^\top Q_1^\top Y) = \left( R_1^\top Q_1^\top Q_1 R_1 \right)^{-1} R_1^\top Q_1^\top Y = \]

\[ = \left( R_1^\top R_1 \right)^{-1} R_1^\top Q_1^\top Y = R_1^{-1} R_1^{-\top} R_1^\top Q_1^\top Y = R_1^{-1} Q_1^\top Y \]

- Solve the following linear system through backward substitution to compute \( \hat{\theta}_{LS} \):

\[ R_1 \hat{\theta}_{LS} = Q_1^\top Y \]
Estimate the parameters $\hat{\theta}_{LS}$ recursively in time.

If there is an estimate $\hat{\theta}_{LS}(k - 1)$ based on data up to time $k - 1$, then $\hat{\theta}_{LS}(k)$ is computed based on a “simple” update of $\hat{\theta}_{LS}(k - 1)$.

No need to record all data up to time $k$ (low memory requirement).

Recursive LS can be easily modified to estimate time-varying parameters.
Recursive linear least squares

\[ \hat{\theta}_{\text{LS}}(k) = \left( \sum_{\ell=1}^{k} \varphi(\ell)\varphi^\top(\ell) \right)^{-1} \sum_{\ell=1}^{k} \varphi(\ell)y(\ell) \]

- \[ P(k) = \left( \sum_{\ell=1}^{k} \varphi(\ell)\varphi^\top(\ell) \right)^{-1}, \quad P^{-1}(k) = P^{-1}(k-1) + \varphi(k)\varphi^\top(k) \]
- \[ \hat{\theta}_{\text{LS}}(k) = P(k) \left( \sum_{\ell=1}^{k-1} \varphi(\ell)y(\ell) + \varphi(k)y(k) \right) \]
- \[ \hat{\theta}_{\text{LS}}(k) = P(k) \left( P^{-1}(k-1)\hat{\theta}_{\text{LS}}(k-1) + \varphi(k)y(k) \right) \]
- \[ \hat{\theta}_{\text{LS}}(k) = \hat{\theta}_{\text{LS}}(k-1) + P(k)\varphi(k) \left( y(k) - \varphi^\top(k)\hat{\theta}_{\text{LS}}(k-1) \right)_{K(k)} \]
- \[ \hat{\theta}_{\text{LS}}(k) = \hat{\theta}_{\text{LS}}(k-1) + K(k)\varepsilon(k) \]
Recursive linear least squares

\[ \hat{\theta}_{LS}(k) = \left( \sum_{\ell=1}^{k} \varphi(\ell)\varphi^\top(\ell) \right)^{-1} \sum_{\ell=1}^{k} \varphi(\ell)y(\ell) \]

- \[ P(k) = \left( \sum_{\ell=1}^{k} \varphi(\ell)\varphi^\top(\ell) \right)^{-1}, \quad P^{-1}(k) = P^{-1}(k-1) + \varphi(k)\varphi^\top(k) \]
- \[ \hat{\theta}_{LS}(k) = P(k) \left( \sum_{\ell=1}^{k-1} \varphi(\ell)y(\ell) + \varphi(k)y(k) \right) \]
- \[ \hat{\theta}_{LS}(k) = P(k) \left( P^{-1}(k-1)\hat{\theta}_{LS}(k-1) + \varphi(k)y(k) \right) \]
- \[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k-1) + P(k)\varphi(k) \left( \frac{y(k) - \varphi^\top(k)\hat{\theta}_{LS}(k-1)}{K(k)} \right) \]
- \[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k-1) + K(k)\varepsilon(k) \]

- \[ K(k): \text{gain} \]
- \[ \varepsilon(k): \text{error in the prediction of } y(k) \text{ based on } \hat{\theta}_{LS}(k-1) \]
Recursive linear least squares

\[ \hat{\theta}_{LS}(k) = \left( \sum_{\ell=1}^{k} \varphi(\ell)\varphi^\top(\ell) \right)^{-1} \sum_{\ell=1}^{k} \varphi(\ell)y(\ell) \]

\[ P(k) = \left( \sum_{\ell=1}^{k} \varphi(\ell)\varphi^\top(\ell) \right)^{-1}, \quad P^{-1}(k) = P^{-1}(k-1) + \varphi(k)\varphi^\top(k) \]

If the prediction error is “small”, the estimate \( \hat{\theta}_{LS}(k-1) \) is “good” and should not be modified “very much”

\[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k-1) + P(k)\varphi(k) \left( y(k) - \varphi^\top(k)\hat{\theta}_{LS}(k-1) \right) \]

\[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k-1) + K(k)\varepsilon(k) \]

- \( K(k) \): gain
- \( \varepsilon(k) \): error in the prediction of \( y(k) \) based on \( \hat{\theta}_{LS}(k - 1) \)
Recursive linear least squares

\[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k - 1) + P(k)\varphi(k) \left( y(k) - \varphi^\top(k)\hat{\theta}_{LS}(k - 1) \right) \]

\[ P^{-1}(k) = P^{-1}(k - 1) + \varphi(k)\varphi^\top(k) \]
Recursive linear least squares

\[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k - 1) + P(k)\varphi(k) \left( y(k) - \varphi^\top(k)\hat{\theta}_{LS}(k - 1) \right) \]

\[ P^{-1}(k) = P^{-1}(k - 1) + \varphi(k)\varphi^\top(k) \]

\( P^{-1}(k) \) can be easily updated
Recursive linear least squares

\[
\hat{\theta}_{\text{LS}}(k) = \hat{\theta}_{\text{LS}}(k - 1) + P(k)\varphi(k) \left( y(k) - \varphi^\top(k)\hat{\theta}_{\text{LS}}(k - 1) \right)
\]

\[
P^{-1}(k) = P^{-1}(k - 1) + \varphi(k)\varphi^\top(k)
\]

\[P^{-1}(k)\] can be easily updated

Updating \(P(k)\) requires to invert \(P^{-1}(k)\) at each time instant (time consuming)
Recursive linear least squares

\[ \hat{\theta}_{LS}(k) = \hat{\theta}_{LS}(k - 1) + P(k)\varphi(k) \left( y(k) - \varphi^\top(k)\hat{\theta}_{LS}(k - 1) \right) \]

\[ P^{-1}(k) = P^{-1}(k - 1) + \varphi(k)\varphi^\top(k) \]

\( P^{-1}(k) \) can be easily updated

Updating \( P(k) \) requires to invert \( P^{-1}(k) \) at each time instant (time consuming)

Solution

\[ P(k) = \left[ P^{-1}(k) \right]^{-1} = \left[ P^{-1}(k - 1) + \varphi(k)\varphi^\top(k) \right]^{-1} \]

From Matrix Inversion Lemma:

\[ P(k) = P(k - 1) - \frac{P(k - 1)\varphi(k)\varphi^\top(k)P(k - 1)}{1 + \varphi^\top(k)P(k - 1)\varphi(k)} \]
Recursive linear LS for real-time identification

- Identify (slowly) time-varying parameters (due to slow time-variation of the process)
- Useful for adaptive control
- Introduce forgetting factor $0 < \lambda \leq 1$ in the cost function:

$$\hat{\theta}(k) = \arg \min_{\theta} \sum_{\ell=1}^{k} \lambda^{k-\ell} \left( y(\ell) - \varphi^\top(\ell)\theta \right)^2$$
Recursive linear LS for real-time identification

- Identify (slowly) time-varying parameters (due to slow time-variation of the process)
- Useful for adaptive control
- Introduce forgetting factor $0 < \lambda \leq 1$ in the cost function:

$$\hat{\theta}(k) = \arg \min_\theta \sum_{\ell=1}^{k} \lambda^{k-\ell} \left( y(\ell) - \varphi^\top(\ell) \theta \right)^2$$

Decrease $\lambda$ to forget information on past data faster
Identify (slowly) time-varying parameters (due to slow time-variation of the process)

Useful for adaptive control

Introduce forgetting factor $0 < \lambda \leq 1$ in the cost function:

$$\hat{\theta}(k) = \arg\min_{\theta} \sum_{\ell=1}^{k} \lambda^{k-\ell} \left( y(\ell) - \varphi^T(\ell)\theta \right)^2$$

Decrease $\lambda$ to forget information on past data faster
Use a Finite Impulse Response (FIR) model to describe the dynamical system $S$ to be identified:

$$\mathcal{M} : \hat{y}(k, g) = \sum_{\ell=0}^{M} g(\ell) u(k - \ell)$$
Use a Finite Impulse Response (FIR) model to describe the dynamical system $S$ to be identified:

$$
\mathcal{M} : \hat{y}(k, g) = \sum_{\ell=0}^{M} g(\ell)u(k - \ell) = \varphi^\top(k)g
$$

$$
\varphi(k) = [u(k) \ u(k - 1) \ \cdots \ u(k - M)]^\top, \quad g = [g(0) \ g(1) \ \cdots g(M)]^\top
$$
Use a Finite Impulse Response (FIR) model to describe the dynamical system $S$ to be identified:

$$M : \hat{y}(k, g) = \sum_{\ell=0}^{M} g(\ell) u(k - \ell) = \varphi^\top(k) g$$

$$\varphi(k) = [u(k) \ u(k-1) \ \cdots \ u(k-M)]^\top, \quad g = [g(0) \ g(1) \ \cdots \ g(M)]^\top$$

Good approximation if $M$ is "large" and $S$ is BIBO stable (which implies $\lim_{\ell \to \infty} |g(\ell)| = 0$)
Estimate of FIR models through LS

Use a Finite Impulse Response (FIR) model to describe the dynamical system $S$ to be identified:

$$
\mathcal{M} : \hat{y}(k, g) = \sum_{\ell=0}^{M} g(\ell) u(k - \ell) = \varphi^\top(k) g
$$

$$
\varphi(k) = [u(k) \ u(k-1) \ \cdots \ u(k-M)]^\top, \quad g = [g(0) \ g(1) \ \cdots \ g(M)]^\top
$$

Good approximation if $M$ is “large” and $S$ is BIBO stable (which implies $\lim_{\ell\to\infty} |g(\ell)| = 0$)

Collect $N \gg M$ observations of the pairs $\{u(k), y(k)\}_{k=1}^{N}$
Use a Finite Impulse Response (FIR) model to describe the dynamical system $S$ to be identified:

$$
\mathcal{M}: \hat{y}(k, g) = \sum_{\ell=0}^{M} g(\ell) u(k - \ell) = \varphi^\top(k) g
$$

$$
\varphi(k) = [u(k) \ u(k-1) \ \cdots \ u(k-M)]^\top, \quad g = [g(0) \ g(1) \ \cdots \ g(M)]^\top
$$

Good approximation if $M$ is “large” and $S$ is BIBO stable (which implies $\lim_{\ell \to \infty} |g(\ell)| = 0$)

Collect $N \gg M$ observations of the pairs $\{u(k), y(k)\}_{k=1}^N$

LS estimate of $g$:

$$
\hat{g} = \left(\Phi^\top \Phi\right)^{-1} \Phi^\top Y
$$
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

Assume that the “true” system \( S \) is described by

\[
S : y(k) = \sum_{\ell=0}^{M} g_o(\ell)u(k - \ell) + v_o(k) = \varphi^\top(k)g_o + v_o(k)
\]
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- Assume that the “true” system \( S \) is described by

\[ S : y(k) = \sum_{\ell=0}^{M} g_\circ(\ell) u(k - \ell) + v_\circ(k) = \varphi^\top(k) g_\circ + v_\circ(k) \]

- \( v_\circ(k) \): zero-mean quasi-stationary noise independent of \( u \)
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- Assume that the “true” system \( S \) is described by

\[
S : y(k) = \sum_{\ell=0}^{M} g_o(\ell)u(k - \ell) + v_o(k) = \varphi^\top(k)g_o + v_o(k)
\]

- \( v_o(k) \): zero-mean quasi-stationary noise independent of \( u \)

Is \( \hat{g} \) a consistent estimate of \( g_o \)?

Does \( \lim_{N \to \infty} \hat{g} = g_o \)?
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

Assume that the “true” system \( S \) is described by

\[ S : \ y(k) = \sum_{\ell=0}^{M} g_o(\ell)u(k - \ell) + v_o(k) = \varphi^\top(k)g_o + v_o(k) \]

\( v_o(k) \): zero-mean quasi-stationary noise independent of \( u \)

Is \( \hat{g} \) a consistent estimate of \( g_o \)?

Does \( \lim_{N \to \infty} \hat{g} = g_o \)?

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top \mathbf{Y} = \]
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

Assume that the “true” system \( S \) is described by

\[ S : y(k) = \sum_{\ell=0}^{M} g_0(\ell)u(k - \ell) + v_o(k) = \varphi^\top(k)g_o + v_o(k) \]

\( v_o(k) \): zero-mean quasi-stationary noise independent of \( u \)

Is \( \hat{g} \) a consistent estimate of \( g_o \)?

Does \( \lim_{N \to \infty} \hat{g} = g_o \)?

\[ \hat{g} = \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^\top(k) \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) = \]
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \]

- Assume that the “true” system \( S \) is described by

\[
S : y(k) = \sum_{\ell=0}^{M} g_o(\ell) u(k - \ell) + v_o(k) = \varphi^\top(k) g_o + v_o(k)
\]

- \( v_o(k) \): zero-mean quasi-stationary noise independent of \( u \)

Is \( \hat{g} \) a consistent estimate of \( g_o \)?

Does \( \lim_{N \to \infty} \hat{g} = g_o \)?

\[
\hat{g} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) y(k) = \\
\left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \left( \varphi^\top(k) g_o + v_o(k) \right) =
\]
Estimate of FIR models through LS

\[ \hat{g} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \]

- Assume that the “true” system \( S \) is described by

\[ S : y(k) = \sum_{\ell=0}^{M} g_o(\ell) u(k - \ell) + v_o(k) = \varphi^T(k) g_o + v_o(k) \]

- \( v_o(k) \): zero-mean quasi-stationary noise independent of \( u \)

Is \( \hat{g} \) a consistent estimate of \( g_o \)?

Does \( \lim_{N \to \infty} \hat{g} = g_o \)?

\[ \hat{g} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) y(k) = \]

\[ \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \left( \varphi^T(k) g_o + v_o(k) \right) = \]

\[ g_o + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) v_o(k) \]
Estimate of FIR models through LS

\[ \hat{g} = g_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_0(k) \]
Estimate of FIR models through LS

\[ \hat{g} = g_o + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_o(k) \]

\[ [R(N)]_{ij} = \left[ \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} u(k - i + 1)u(k - j + 1) \]
Estimate of FIR models through LS

\[ \hat{g} = g_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_0(k) \]

- \[ [R(N)]_{ij} = \left[ \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} u(k - i + 1)u(k - j + 1) \]

- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} u(k)u(k - \tau) = \bar{E}[u(k)u(k - \tau)] = R_u(\tau) \)
Estimate of FIR models through LS

\[ \hat{g} = g_o + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_o(k) \]

- \( [R(N)]_{ij} = \left[ \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} u(k - i + 1)u(k - j + 1) \)

- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} u(k)u(k - \tau) = \mathbb{E} [u(k)u(k - \tau)] = R_u(\tau) \)

- Thus, \( \lim_{N \to \infty} R(N) = R^* = \mathbb{E} \left[ \varphi(k)\varphi^\top(K) \right] \)
Estimate of FIR models through LS

\[ \hat{g} = g_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_o(k) \]

- \[[R(N)]_{ij} = \left[ \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} u(k - i + 1)u(k - j + 1)\]

- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} u(k)u(k - \tau) = \overline{E}[u(k)u(k - \tau)] = R_u(\tau) \)

- Thus, \( \lim_{N \to \infty} R(N) = R^* = \overline{E}\left[\varphi(k)\varphi^\top(K)\right] \)

- \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_o(k) = \overline{E}[\varphi(k)v_o(k)] = 0 \), since \( u(k) \) is independent of \( v_o(k) \) and \( \overline{E}[v_o(k)] = 0 \).
Estimate of FIR models through LS

\[
\hat{g} = g_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_o(k)
\]

- \([R(N)]_{ij} = \left[ \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} u(k - i + 1)u(k - j + 1)\]

- If \(u\) is quasi-stationary, \(\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} u(k)u(k - \tau) = \mathbb{E} [u(k)u(k - \tau)] = R_u(\tau)\)

Thus, \(\lim_{N \to \infty} R(N) = R^* = \mathbb{E} \left[ \varphi(k)\varphi^T(K) \right]\)

- \(\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v_o(k) = \mathbb{E} [\varphi(k)v_o(k)] = 0\), since \(u(k)\) is independent of \(v_o(k)\) and \(\mathbb{E} [v_o(k)] = 0\).

- \(\lim_{N \to \infty} \hat{g} = g_0 + R^* \mathbb{E} [\varphi(k)v_o(k)] = g_0\)
Estimate of FIR models through LS

\[
\hat{g} = g_o + \left( \frac{1}{N} \sum_{k=1}^{N} \phi(k)\phi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \phi(k)v_o(k)
\]

- \([R(N)]_{ij} = \left[ \frac{1}{N} \sum_{k=1}^{N} \phi(k)\phi^\top(k) \right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} u(k-i+1)u(k-j+1)\]

- If \(u\) is quasi-stationary, \(\lim \frac{1}{N} \sum_{k=1}^{N} u(k)u(k-\tau) = \overline{E} \left[ u(k)u(k-\tau) \right] = R_u(\tau)\)

- Thus, \(\lim_{N \rightarrow \infty} R(N) = R^* = \overline{E} \left[ \phi(k)\phi^\top(K) \right]\)

- \(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{N} \phi(k)v_o(k) = \overline{E} \left[ \phi(k)v_o(k) \right] = 0\), since \(u(k)\) is independent of \(v_o(k)\) and \(\overline{E} [v_o(k)] = 0\).

- \(\lim_{N \rightarrow \infty} \hat{g} = g_o + R^*\overline{E} [\phi(k)v_o(k)] = g_o\)

- Thus, \(\hat{g}\) is a consistent estimate of \(g_o\)
Estimate of FIR models through LS: example

\( N = 300 \)

**Pulse response**

- true
- estimate

**output**

\[
SNR = 10 \log_{10} \left( \frac{\sum_{k=1}^{N} y^2(k)}{\sum_{k=1}^{N} v^2_0(k)} \right) = 6 \text{ db}
\]

\[
BFR = 1 - \frac{\sum_{k=1}^{N_{v}} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N_{v}} (y(k) - \bar{y})^2} = 72\%
\]
Estimate of FIR models through LS: example

\( N = 500 \)

\[
\text{SNR} = 10 \log_{10} \left( \frac{\sum_{k=1}^{N} y^2(k)}{\sum_{k=1}^{N} v_o^2(k)} \right) = 6 \text{ db}
\]

\[
BFR = 1 - \frac{\sum_{k=1}^{N_v} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N_v} (y(k) - \bar{y})^2} = 91\%
\]
Estimate of FIR models through LS: example

\[ N = 1000 \]

\[ \text{Pulse response} \]

\[ \text{output} \]

\[ \text{SNR} = 10 \log_{10} \left( \frac{\sum_{k=1}^{N} y^2(k)}{\sum_{k=1}^{N} \nu_o^2(k)} \right) = 6 \text{ db} \]

\[ BFR = 1 - \frac{\sum_{k=1}^{N_y} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N_y} (y(k) - \bar{y})^2} = 96\% \]
Estimate of FIR models through LS: example

$N = 5000$

\[ SNR = 10 \log_{10} \left( \frac{\sum_{k=1}^{N} y^2(k)}{\sum_{k=1}^{N} v^2_o(k)} \right) = 6 \text{ db} \]

\[ BFR = 1 - \frac{\sum_{k=1}^{N} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N} (y(k) - \bar{y})^2} = 99\% \]
Estimate of FIR models through LS: example

$N = 10000$

$$SNR = 10 \log_{10} \left( \frac{\sum_{k=1}^{N} y^2(k)}{\sum_{k=1}^{N} v^2_0(k)} \right) = 6 \text{ db}$$

Pulse response

output

$$BFR = 1 - \frac{\sum_{k=1}^{N_v} (y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N_v} (y(k) - \bar{y})^2} = 99.7\%$$
Estimate of ARX models through LS

\[ y(k) = G(q)u(k) + H(q)e(k) \]

\[ G(q) = \begin{bmatrix} b_1 q^1 & \cdots & b_n q^n \end{bmatrix} \]

\[ A(q) = 1 + a_1 q^1 + \cdots + a_n q^n \]

\[ H(q) = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \]

Corresponding input/output relationship

\[ y(k) = a_1 y(k-1) + \cdots + a_n y(k-n) + b_1 u(k-1) + \cdots + b_n u(k-n) + e(k) \]

Unknown parameter vector:

\[ \theta = [a_1 : : a_n : : b_1 : : b_n] ^\top \]

Regressor vector:

\[ \varphi = [y(k-1) : : y(k-n) u(k-1) : : u(k-n)] ^\top \]

Output representation:

\[ y(k) = \varphi ^\top (k) + e(k) \]
Estimate of ARX models through LS

\[ y(k) = G(q)u(k) + H(q)e(k) \]

- AutoRegressive with eXogenous input (ARX) model structure:

- \( G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}} \)

- \( H(q) = \frac{1}{A(q)} = \frac{1}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}} \)
Estimate of ARX models through LS

\[ y(k) = G(q)u(k) + H(q)e(k) \]

- AutoRegressive with eXogenous input (ARX) model structure:

\[
G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}}
\]

\[
H(q) = \frac{1}{A(q)} = \frac{1}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}}
\]

- Corresponding input/output relationship

\[
y(k) = -a_1 y(k - 1) - \cdots - a_{n_a} y(k - n_a) + b_1 u(k - 1) + \cdots + b_{n_b} u(k - n_b) + e(k)
\]
Estimate of ARX models through LS

\[ y(k) = G(q)u(k) + H(q)e(k) \]

- AutoRegressive with eXogenous input (ARX) model structure:
  \[
  G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}} \\
  H(q) = \frac{1}{A(q)} = \frac{1}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}}
  \]

- Corresponding input/output relationship
  \[ y(k) = -a_1 y(k-1) - \cdots - a_{na} y(k-n_a) + b_1 u(k-1) + \cdots + b_{nb} u(k-n_b) + e(k) \]

- Unknown parameter vector: \[ \theta = [a_1 \ldots a_{na} \ b_1 \ldots b_{nb}]^\top \]

- Regressor vector: \[ \varphi = [-y(k-1) \ldots -y(k-n_a) \ u(k-1) \ldots u(k-n_b)]^\top \]

- Output representation: \[ y(k) = \varphi^\top(k) \theta + e(k) \]
Output representation: \( y(k) = \varphi^\top(k)\theta + e(k) \)
Output representation: \( y(k) = \varphi^T(k)\theta + e(k) \)

Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
Estimate of ARX models through LS

- Output representation: \( y(k) = \varphi^T(k)\theta + e(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = (\Phi^T \Phi)^{-1} \Phi^T Y \)
Estimate of ARX models through LS

- Output representation: \( y(k) = \varphi^\top(k)\theta + e(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^\top(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg\min_{\theta} \| Y - \Phi\theta \|^2 = \left( \Phi^\top\Phi \right)^{-1} \Phi^\top Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^\top(k)\theta_0 + e(k) \)
Estimate of ARX models through LS

- Output representation: \( y(k) = \phi^T(k)\theta + e(k) \)
- Model structure: \( M : \hat{y}(k, \theta) = \phi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg \min_{\theta} \| Y - \Phi\theta \|^2 = \left( \Phi^T\Phi \right)^{-1} \Phi^T Y \)

Assume the “true” system is described by \( S : y(k) = \phi^T(k)\theta_0 + e(k) \)

Is \( \hat{\theta}_{LS} \) a consistent estimate of \( \theta_0 \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{LS} = \theta_0 \)?
Estimate of ARX models through LS

- Output representation: $y(k) = \varphi^T(k)\theta + e(k)$
- Model structure: $\mathcal{M}: \hat{y}(k, \theta) = \varphi^T(k)\theta$ (linear regression)
- LS estimate: $\hat{\theta}_{LS} = \arg\min_{\theta} \| Y - \Phi \theta \|^2 = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y$
- Assume the “true” system is described by $S: y(k) = \varphi^T(k)\theta_0 + e(k)$

Is $\hat{\theta}_{LS}$ a consistent estimate of $\theta_0$?

Does $\lim_{N \to \infty} \hat{\theta}_{LS} = \theta_0$?

$\hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y =$
Estimate of ARX models through LS

- Output representation: \( y(k) = \varphi^T(k)\theta + e(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{\text{LS}} = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \)

Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_0 + e(k) \)

**Is \( \hat{\theta}_{\text{LS}} \) a consistent estimate of \( \theta_0 \)?**

Does \( \lim_{N \to \infty} \hat{\theta}_{\text{LS}} = \theta_0 \)?

\[
\hat{\theta}_{\text{LS}} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) =
\]
Estimate of ARX models through LS

- **Output representation:** $y(k) = \varphi^T(k)\theta + e(k)$
- **Model structure:** $\mathcal{M}: \hat{y}(k, \theta) = \varphi^T(k)\theta$ (linear regression)
- **LS estimate:** $\hat{\theta}_{LS} = \arg\min_\theta \| Y - \Phi\theta \|^2 = \left( \Phi^T\Phi \right)^{-1} \Phi^T Y$
- **Assume the “true” system is described by $S: y(k) = \varphi^T(k)\theta_o + e(k)$**

**Is $\hat{\theta}_{LS}$ a consistent estimate of $\theta_o$?**

Where $\theta_o$ is the true parameter vector of the system.

Does $\lim_{N\to\infty} \hat{\theta}_{LS} = \theta_o$?

\[
\hat{\theta}_{LS} = \left( \Phi^T\Phi \right)^{-1} \Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) = \\
\left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \left( \varphi^T(k)\theta_o + e(k) \right) =
\]
Estimate of ARX models through LS

- Output representation: \( y(k) = \varphi^T(k)\theta + e(k) \)
- Model structure: \( M: \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg\min_{\theta} \| Y - \Phi\theta \|^2 = (\Phi^T\Phi)^{-1}\Phi^T Y \)
- Assume the “true” system is described by \( S: y(k) = \varphi^T(k)\theta_o + e(k) \)

Is \( \hat{\theta}_{LS} \) a consistent estimate of \( \theta_o \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{LS} = \theta_o \)?

\[
\hat{\theta}_{LS} = (\Phi^T\Phi)^{-1}\Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) = \\
\frac{1}{N} \sum_{k=1}^{N} \varphi(k)\left( \varphi^T(k)\theta_o + e(k) \right) = \\
\theta_o + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k)
\]
Estimate of ARX models through LS

\[ \hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) \]
Estimate of ARX models through LS

\[ \hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates
Estimate of ARX models through LS

\[
\hat{\theta}_{\text{LS}} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k)
\]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates
- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} R(N) = R^* \)
Estimate of ARX models through LS

\[ \hat{\theta}_{LS} = \theta_o + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates.
- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} R(N) = R^* \)
- \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) = \mathbb{E} [\varphi(k)e(k)] = \mathbb{E} [\varphi(k)] \mathbb{E} [e(k)] = 0 \), since \( \varphi(k) \) is independent of \( e(k) \) and \( \mathbb{E} [e(k)] = 0 \).
Estimate of ARX models through LS

\[
\hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k)
\]

- \(R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k)\) is filled out with the estimate of auto/cross covariance function estimates.
- If \(u\) is quasi-stationary, \(\lim_{N \to \infty} R(N) = R^*\)
- \(\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) = \mathbb{E}[\varphi(k)e(k)] = \mathbb{E}[\varphi(k)] \mathbb{E}[e(k)] = 0\), since \(\varphi(k)\) is independent of \(e(k)\) and \(\mathbb{E}[e(k)] = 0\).
- \(\lim_{N \to \infty} \hat{\theta}_{LS} = \theta_0 + R^*\mathbb{E}[\varphi(k)e(k)] = \theta_0\)
Estimate of ARX models through LS

\[ \hat{\theta}_{\text{LS}} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates.
- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} R(N) = R^* \).
- \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)e(k) = \mathbb{E}[\varphi(k)e(k)] = \mathbb{E}[\varphi(k)]\mathbb{E}[e(k)] = 0 \), since \( \varphi(k) \) is independent of \( e(k) \) and \( \mathbb{E}[e(k)] = 0 \).
- \( \lim_{N \to \infty} \hat{\theta}_{\text{LS}} = \theta_0 + R^*\mathbb{E}[\varphi(k)e(k)] = \theta_0 \).
- Thus, \( \hat{\theta}_{\text{LS}} \) is a consistent estimate of \( \theta_0 \).
#### Linear Parameter-Varying (LPV) systems

- Linear relationship between inputs and outputs:

\[ y(k) = G(q^{-1}, p(k))u(k) + v(k) \]

- The input/output relationship changes over time according to a **measurable** signal \( p \) (called **scheduling signal**)

*Figure provided by R. Tóth*
Linear Parameter-Varying (LPV) systems

- Linear relationship between inputs and outputs:

\[ y(k) = G(q^{-1}, p(k))u(k) + v(k) \]

- The input/output relationship changes over time according to a measurable signal \( p \) (called scheduling signal)

- The scheduling signal can be, for instance, an external variable used to describe different operating conditions of the system (e.g., temperature, space coordinates)
Estimate of LPV-ARX models through LS

Linear Parameter-Varying (LPV) systems

- Linear relationship between inputs and outputs:
  \[ y(k) = G(q^{-1}, p(k))u(k) + v(k) \]

- The input/output relationship changes over time according to a measurable signal \( p \) (called scheduling signal)

- The scheduling signal can be, for instance, an external variable used to describe different operating conditions of the system (e.g., temperature, space coordinates)

Figure provided by R. Tóth

LPV-ARX models

- \[ A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k), \quad e \text{ white} \]
Estimate of LPV-ARX models through LS

**Linear Parameter-Varying (LPV) systems**

- Linear relationship between inputs and outputs:

\[ y(k) = G(q^{-1}, p(k))u(k) + v(k) \]

- The input/output relationship changes over time according to a measurable signal \( p \) (called scheduling signal)

- The scheduling signal can be, for instance, an external variable used to describe different operating conditions of the system (e.g., temperature, space coordinates)

**LPV-ARX models**

- \( A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k) \), \( e \) white

- \( A(q^{-1}, p(k)) = 1 + \sum_{i=1}^{n_a} a_i(p(k))q^{-i} \), \( B(q^{-1}, p(k)) = \sum_{j=1}^{n_b} b_j(p(k))q^{-j} \)
Estimate of LPV-ARX models through LS

Linear Parameter-Varying (LPV) systems

- Linear relationship between inputs and outputs:
  \[ y(k) = G(q^{-1}, p(k))u(k) + v(k) \]
- The input/output relationship changes over time according to a measurable signal \( p \) (called scheduling signal)
- The scheduling signal can be, for instance, an external variable used to describe different operating conditions of the system (e.g., temperature, space coordinates)

LPV-ARX models

- \( A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k), \quad e \text{ white} \)
- \( A(q^{-1}, p(k)) = 1 + \sum_{i=1}^{n_a} a_i(p(k))q^{-i}, \quad B(q^{-1}, p(k)) = \sum_{j=1}^{n_b} b_j(p(k))q^{-j} \)
- \( a_i(p(k)) \) and \( b_j(p(k)) \) are a-priori parametrized functions of \( p(k) \) (e.g., polynomials):
  \[ a_i(p(k)) = a_{i,0} + \sum_{l=1}^{n_i} a_{i,l}p^l(k), \quad b_j(p(k)) = b_{j,0} + \sum_{l=1}^{n_j} b_{j,l}p^l(k) \]
Estimate of LPV-ARX models through LS

LPV-ARX models: example

- \( A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k), \quad e \text{ white} \)
- Example:

\[
y(k) = - [a_{1,0} + a_{1,1}p(k)] y(k - 1) + [b_{1,0} + b_{1,1}p(k)] u(k - 1) + e(k)
\]
Estimate of LPV-ARX models through LS

LPV-ARX models: example

- $A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k), \quad e \text{ white}$

Example:

\[
y(k) = -[a_{1,0} + a_{1,1}p(k)]y(k-1) + [b_{1,0} + b_{1,1}p(k)]u(k-1) + e(k)
\]

\[
y(k) = \begin{bmatrix} -y(k-1) & -p(k)y(k-1) & u(k-1) & p(k)u(k-1) \end{bmatrix} \begin{bmatrix} a_{1,0} \\ a_{1,1} \\ b_{1,0} \\ b_{1,1} \end{bmatrix} + e(k)
\]

\[\begin{bmatrix} a_{1,0} \\ a_{1,1} \\ b_{1,0} \\ b_{1,1} \end{bmatrix} = \phi^\top(k) \theta \]
Estimate of LPV-ARX models through LS

LPV-ARX models: example

- \( A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k), \quad e \text{ white} \)

Example:

\[
y(k) = -[a_{1,0} + a_{1,1}p(k)]y(k-1) + [b_{1,0} + b_{1,1}p(k)]u(k-1) + e(k)
\]

\[
y(k) = \begin{bmatrix} -y(k-1) & -p(k)y(k-1) & u(k-1) & p(k)u(k-1) \end{bmatrix} \begin{bmatrix} a_{1,0} \\ a_{1,1} \\ b_{1,0} \\ b_{1,1} \end{bmatrix} + e(k)
\]

- Estimate \( \theta \) through least-squares
Estimate of LPV-ARX models through LS

LPV-ARX models: example

- $A(q^{-1}, p(k))y(k) = B(q^{-1}, p(k))u(k) + e(k)$, $e$ white

Example:

$$y(k) = - [a_{1,0} + a_{1,1}p(k)] y(k - 1) + [b_{1,0} + b_{1,1}p(k)] u(k - 1) + e(k)$$

$$y(k) = \begin{bmatrix} y(k - 1) & -p(k)y(k - 1) & u(k - 1) & p(k)u(k - 1) \end{bmatrix} \varphi^T(k) + e(k)$$

- Estimate $\theta$ through least-squares
- Consistency is guaranteed if $e$ is white
Inconsistency of LS: OE case

$$y(k) = G(q)u(k) + H(q)e(k)$$
Inconsistency of LS: OE case

\[ y(k) = G(q)u(k) + H(q)e(k) \]

Output Error (OE) model structure:

\[
G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}}
\]

\[ H(q) = 1 \]

\[ y \]

[Diagram of a system with inputs, outputs, and error signals]
\[ y(k) = G(q)u(k) + H(q)e(k) \]

- **Output Error (OE) model structure:**
  \[
  G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}} \\
  H(q) = 1
  \]

- **Corresponding input/output relationship**
  \[
  y_o(k) = -a_1 y_o(k - 1) - \cdots - a_{na} y_o(k - n_a) + b_1 u(k - 1) + \cdots + b_{nb} u(k - n_b) \\
  y(k) = y_o(k) + e(k)
  \]
Inconsistency of LS: OE case

\[ y(k) = G(q)u(k) + H(q)e(k) \]

- Output Error (OE) model structure:

  \[ G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}} \]

  \[ H(q) = 1 \]

- Corresponding input/output relationship

  \[ y_o(k) = -a_1 y_o(k - 1) - \cdots - a_{na} y_o(k - n_a) + b_1 u(k - 1) + \cdots + b_{nb} u(k - n_b) \]

  \[ y(k) = y_o(k) + e(k) \]

- Unknown parameter vector: \( \theta = [a_1 \ldots a_{na} \ b_1 \ldots b_{nb}]^T \)

- Regressor vector: \( \varphi = [-y(k-1) \ldots -y(k-n_a) \ u(k-1) \ldots u(k-n_b)]^T \)

- Output representation: \( y(k) = \varphi^T(k) \theta + e(k) + a_1 e(k-1) + \cdots + a_{na} e(k - n_a) \)

\( v(k) \) is not white!
Inconsistency of LS: OE case

\[ y(k) = G(q)u(k) + H(q)e(k) \]

Output Error (OE) model structure:

\[ G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + \cdots + b_{nb} q^{-nb}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}} \]

\[ H(q) = 1 \]

Corresponding input/output relationship

\[ y_o(k) = -a_1 y_o(k - 1) - \cdots - a_{na} y_o(k - n_a) + b_1 u(k - 1) + \cdots + b_{nb} u(k - n_b) \]

\[ y(k) = y_o(k) + e(k) \]

Unknown parameter vector: \( \theta = [a_1 \ldots a_{na} b_1 \ldots b_{nb}]^\top \)

Regressor vector: \( \varphi = [-y(k - 1) \ldots - y(k - n_a) u(k - 1) \ldots u(k - n_b)]^\top \)

Output representation: \( y(k) = \varphi^\top(k) \theta + e(k) + a_1 e(k - 1) + \cdots + a_{na} e(k - n_a) \)

\( \nu(k) \) is not white!
Output representation: \( y(k) = \varphi^T(k) \theta + \nu(k) \)
Output representation: $y(k) = \varphi^T(k)\theta + v(k)$

Model structure: $M: \hat{y}(k, \theta) = \varphi^T(k)\theta$ (linear regression)
Inconsistency of LS: OE case

- Output representation: \( y(k) = \varphi^\top(k)\theta + v(k) \)
- Model structure: \( M : \hat{y}(k, \theta) = \varphi^\top(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg\min_{\theta} \| V \|^2 = \arg\min_{\theta} \| Y - \Phi\theta \|^2 = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y \)
Inconsistency of LS: OE case

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( M : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg \min_{\theta} \| V \|^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_0 + \nu(k) \)
Inconsistency of LS: OE case

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)

- Model structure: \( M: \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)

- LS estimate: \( \hat{\theta}_{LS} = \arg\min_{\theta} \| V \|^2 = \arg\min_{\theta} \| Y - \Phi\theta \|^2 = \left( \Phi^T\Phi \right)^{-1} \Phi^T Y \)

- Assume the “true” system is described by \( S: y(k) = \varphi^T(k)\theta_o + \nu(k) \)

**Is \( \hat{\theta}_{LS} \) a consistent estimate of \( \theta_o \)?**

**Does \( \lim_{N \to \infty} \hat{\theta}_{LS} = \theta_o \)?**
Inconsistency of LS: OE case

- Output representation: $y(k) = \varphi^T(k)\theta + v(k)$
- Model structure: $\mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta$ (linear regression)
- LS estimate: $\hat{\theta}_{LS} = \arg \min_{\theta} \| V \|^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = (\Phi^T \Phi)^{-1} \Phi^T Y$
- Assume the “true” system is described by $S : y(k) = \varphi^T(k)\theta_o + v(k)$

Is $\hat{\theta}_{LS}$ a consistent estimate of $\theta_o$?

Does $\lim_{N \to \infty} \hat{\theta}_{LS} = \theta_o$?

$\hat{\theta}_{LS} = (\Phi^T \Phi)^{-1} \Phi^T Y =$
Inconsistency of LS: OE case

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg \min_{\theta} \| V \|^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_o + \nu(k) \)

**Is \( \hat{\theta}_{LS} \) a consistent estimate of \( \theta_o \)?**

Does \( \lim_{N \to \infty} \hat{\theta}_{LS} = \theta_o \)?

\[
\hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) =
\]
Inconsistency of LS: OE case

- Output representation: \( y(k) = \varphi^T(k) \theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k) \theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg \min_{\theta} \| V \|^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k) \theta_0 + \nu(k) \)

Is \( \hat{\theta}_{LS} \) a consistent estimate of \( \theta_0 \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{LS} = \theta_0 \)?

\[
\hat{\theta}_{LS} = \left( \Phi^T \Phi \right)^{-1} \Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) y(k) = \\
\left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \left( \varphi^T(k) \theta_0 + \nu(k) \right)
\]
Inconsistency of LS: OE case

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- LS estimate: \( \hat{\theta}_{LS} = \arg \min_{\theta} \| V \|^2 = \arg \min_{\theta} \| Y - \Phi \theta \|^2 = (\Phi^T \Phi)^{-1} \Phi^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_\circ + \nu(k) \)

Is \( \hat{\theta}_{LS} \) a consistent estimate of \( \theta_\circ \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{LS} = \theta_\circ ? \)

\[
\hat{\theta}_{LS} = (\Phi^T \Phi)^{-1} \Phi^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) =
\]

\[
\left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \left( \varphi^T(k)\theta_\circ + \nu(k) \right) =
\]

\[
\theta_\circ + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\nu(k)
\]
Inconsistency of LS: OE case

\[ \hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v(k) \]
Inconsistency of LS: OE case

\[ \hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates
Inconsistency of LS: OE case

\[ \hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) v(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates
- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} R(N) = R^* \)
Inconsistency of LS: OE case

\[ \hat{\theta}_{LS} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates.
- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} R(N) = R^* \).
- \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)v(k) = \mathbb{E} [\varphi(k)v(k)] \neq 0 \) since \( \varphi(k) \) is correlated with \( v(k) \).
Inconsistency of LS: OE case

\[ \hat{\theta}_{\text{LS}} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\nu(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \) is filled out with the estimate of auto/cross covariance function estimates
- If \( u \) is quasi-stationary, \( \lim_{N \to \infty} R(N) = R^* \)
- \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\nu(k) = \mathbb{E} [\varphi(k)\nu(k)] \neq 0 \) since \( \varphi(k) \) is correlated with \( \nu(k) \)
- Thus, \( \hat{\theta}_{\text{LS}} \) is not a consistent estimate of \( \theta_0 \)
Instrumental Variable Methods
Given a model structure \( A(q^{-1})y(k) = B(q^{-1})u(k) + v(k) \), LS provides a consistent estimate of the “true” system parameters only when \( \{v(k)\} \) is not correlated with the regressor (equivalently, if \( v \) is white).
**Instrumental Variables (IV)**

\[ y(k) = G(q)u(k) + H(q)e(k) \]

- Given a model structure \( A(q^{-1})y(k) = B(q^{-1})u(k) + v(k) \), LS provides a consistent estimate of the “true” system parameters only when \( \{v(k)\} \) is not correlated with the regressor (equivalently, if \( v \) is white).

- Instrumental Variables (IV) methods provide solutions to guarantee consistency also when \( \{v(k)\} \) is correlated with the regressor.
Instrumental Variables Estimate

- Chose a vector \( z(k) \), called instrument, with the same dimension of the regressor \( \varphi(k) \) and such that

  \[ \mathbb{E}[z(k)v(k)] = 0 \quad \text{(i.e., } z(k) \text{ is not correlated with } v(k)) \]

- Modify the LS estimate as follows

  \[
  \hat{\theta}_{IV} = \left( Z^\top \Phi \right)^{-1} Z^\top Y = \left( \sum_{k=1}^{N} z(k)\varphi^\top(k) \right)^{-1} \sum_{k=1}^{N} z(k)y(k)
  \]

  with

  \[
  Z = \begin{bmatrix}
  z^\top(1) \\
  \vdots \\
  z^\top(N)
  \end{bmatrix}
  \]
Instrumental Variables (IV): main idea

\[
\hat{\theta}_{\text{LS}} = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y = \left( \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right)^{-1} \sum_{k=1}^{N} \varphi(k)y(k)
\]

Instrumental Variables Estimate

- Chose a vector \( z(k) \), called instrument, with the same dimension of the regressor \( \varphi(k) \) and such that

  \[
  \mathbb{E} [z(k)\nu(k)] = 0 \quad (\text{i.e., } z(k) \text{ is not correlated with } \nu(k))
  \]

- Modify the LS estimate as follows

  \[
  \hat{\theta}_{\text{IV}} = \left( Z^\top \Phi \right)^{-1} Z^\top Y = \left( \sum_{k=1}^{N} z(k)\varphi^\top(k) \right)^{-1} \sum_{k=1}^{N} z(k)y(k)
  \]

  with

  \[
  \begin{bmatrix}
  z^\top(1) \\
  \vdots \\
  z^\top(N)
  \end{bmatrix} \hat{\theta}_{\text{IV}} = Z^\top Y \rightarrow R\hat{\theta}_{\text{IV}} = Q^\top Z^\top Y, \quad [Q, R] = qr(Z^\top \Phi)
  \]
Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)

Model structure: \( \mathcal{M}: \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
Instrumental Variables: consistency analysis

- Output representation: $y(k) = \varphi^T(k)\theta + \nu(k)$
- Model structure: $\mathcal{M}: \hat{y}(k, \theta) = \varphi^T(k)\theta$ (linear regression)
- IV estimate: $\hat{\theta}_{IV} = \left(Z^T\varphi\right)^{-1}Z^TY$
Instrumental Variables: consistency analysis

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- IV estimate: \( \hat{\theta}_{IV} = \left( Z^T \Phi \right)^{-1} Z^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_\circ + \nu(k) \)
Instrumental Variables: consistency analysis

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- IV estimate: \( \hat{\theta}_{IV} = \left( Z^T \Phi \right)^{-1} Z^T Y \)
- Assume the “true” system is described by \( \mathcal{S} : y(k) = \varphi^T(k)\theta_0 + \nu(k) \)

Is \( \hat{\theta}_{IV} \) a consistent estimate of \( \theta_0 \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{IV} = \theta_0 \)?
Instrumental Variables: consistency analysis

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- IV estimate: \( \hat{\theta}_{IV} = \left( Z^T \Phi \right)^{-1} Z^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_0 + \nu(k) \)

Is \( \hat{\theta}_{IV} \) a consistent estimate of \( \theta_0 \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{IV} = \theta_0 \) ?

\[
\hat{\theta}_{IV} = \left( Z^T \Phi \right)^{-1} Z^T Y =
\]
Instrumental Variables: consistency analysis

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- IV estimate: \( \hat{\theta}_{IV} = \left(Z^T\varphi\right)^{-1}Z^TY \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_0 + \nu(k) \)

Is \( \hat{\theta}_{IV} \) a consistent estimate of \( \theta_0 \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{IV} = \theta_0 \)?

\[
\hat{\theta}_{IV} = \left(Z^T\varphi\right)^{-1}Z^TY = \left(\frac{1}{N}\sum_{k=1}^{N}z(k)\varphi^T(k)\right)^{-1}\frac{1}{N}\sum_{k=1}^{N}z(k)y(k) =
\]
Instrumental Variables: consistency analysis

- Output representation: \( y(k) = \varphi(k)^\top \theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi(k)^\top \theta \) (linear regression)
- IV estimate: \( \hat{\theta}_{IV} = \left( Z^\top \Phi \right)^{-1} Z^\top Y \)

Assume the “true” system is described by \( S : y(k) = \varphi(k)^\top \theta_o + \nu(k) \)

Is \( \hat{\theta}_{IV} \) a consistent estimate of \( \theta_o \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{IV} = \theta_o \)?

\[
\hat{\theta}_{IV} = \left( Z^\top \Phi \right)^{-1} Z^\top Y = \left( \frac{1}{N} \sum_{k=1}^{N} z(k) \varphi(k)^\top \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k) y(k) = \\
\left( \frac{1}{N} \sum_{k=1}^{N} z(k) \varphi(k)^\top \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k) \left( \varphi(k)^\top \theta_o + \nu(k) \right) = 
\]
Instrumental Variables: consistency analysis

- Output representation: \( y(k) = \varphi^T(k)\theta + \nu(k) \)
- Model structure: \( \mathcal{M} : \hat{y}(k, \theta) = \varphi^T(k)\theta \) (linear regression)
- IV estimate: \( \hat{\theta}_{IV} = \left( Z^T \Phi \right)^{-1} Z^T Y \)
- Assume the “true” system is described by \( S : y(k) = \varphi^T(k)\theta_o + \nu(k) \)

Is \( \hat{\theta}_{IV} \) a consistent estimate of \( \theta_o \)?

Does \( \lim_{N \to \infty} \hat{\theta}_{IV} = \theta_o \)?

\[
\hat{\theta}_{IV} = \left( Z^T \Phi \right)^{-1} Z^T Y = \left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k)y(k) = \\
\left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k) \left( \varphi^T(k)\theta_o + \nu(k) \right) = \\
\theta_o + \left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k)\nu(k)
\]
Instrumental Variables: consistency analysis

\[ \hat{\theta}_{IV} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k)v(k) \]
Instrumental Variables: consistency analysis

\[ \hat{\theta}_{\text{IV}} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k)v(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^\top(k) \) is filled out with the estimate of cross covariance function estimates
\[ \hat{\theta}_{IV} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} z(k) \varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k) v(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} z(k) \varphi^T(k) \) is filled out with the estimate of cross covariance function estimates
- Under mild assumptions, \( \lim_{N \to \infty} R(N) = R^* \)
- \( z(k) \) should be correlated with \( \varphi(k) \) (but not with \( v(k) \)) otherwise \( R(N) \) converges to a zero matrix!
\[ \hat{\theta}_{IV} = \theta_o + \left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^\top(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k)v(k) \]

- \( R(N) = \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^\top(k) \) is filled out with the estimate of cross covariance function estimates.
- Under mild assumptions, \( \lim_{N \to \infty} R(N) = R^* \)
- \( z(k) \) should be correlated with \( \varphi(k) \) (but not with \( v(k) \)) otherwise \( R(N) \) converges to a zero matrix!
- \( \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} z(k)v(k) = \overline{E} [z(k)v(k)] = 0 \)
Instrumental Variables: consistency analysis

$$\hat{\theta}_{IV} = \theta_0 + \left( \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^T(k) \right)^{-1} \frac{1}{N} \sum_{k=1}^{N} z(k)\nu(k)$$

- $R(N) = \frac{1}{N} \sum_{k=1}^{N} z(k)\varphi^T(k)$ is filled out with the estimate of cross covariance function estimates
- Under mild assumptions, $\lim_{N \to \infty} R(N) = R^*$
- $z(k)$ should be correlated with $\varphi(k)$ (but not with $\nu(k)$!) otherwise $R(N)$ converges to a zero matrix!
- $\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} z(k)\nu(k) = \overline{E}[z(k)\nu(k)] = 0$
- Thus, $\hat{\theta}_{IV}$ is a consistent estimate of $\theta_0$
The instrument $z(k)$ should be such that $\mathbb{E}[z(k)v(k)] = 0$.
Instrumental Variables: how to choose the instruments?

- The instrument $z(k)$ should be such that $\mathbb{E}[z(k)v(k)] = 0$
- The instrument $z(k)$ should be correlated with $\varphi(k)$
- In order to minimize the variance of the estimate, we would like to choose:

$$z(k) = [-y_o(k - 1) \cdots - y_o(k - n_a) \ u(k - 1) \cdots u(k - n_b)]^T$$
The instrument $z(k)$ should be such that $\mathbb{E}[z(k)v(k)] = 0$

The instrument $z(k)$ should be correlated with $\varphi(k)$

In order to minimize the variance of the estimate, we would like to choose:

$$z(k) = [-y_o(k - 1) \cdots - y_o(k - n_a) \ u(k - 1) \cdots u(k - n_b)]^T$$

Summarizing, $z(k)$ should be as much as possible correlated with the noise-free regressor, but not correlated with the residual $v(k)$. 
The instrument \( z(k) \) should be such that \( \mathbb{E}[z(k)v(k)] = 0 \)

The instrument \( z(k) \) should be correlated with \( \varphi(k) \)

In order to minimize the variance of the estimate, we would like to choose:

\[
z(k) = \begin{bmatrix}
    -y_o(k - 1) \\
    \vdots \\
    -y_o(k - n_a) \\
    u(k - 1) \\
    \vdots \\
    u(k - n_b)
\end{bmatrix}^T
\]

Summarizing, \( z(k) \) should be as much as possible correlated with the noise-free regressor, but not correlated with the residual \( v(k) \).

Choice of the instrument: idea

1. Estimate the model parameters \( \hat{\theta} \) through LS (biased estimate!)
The instrument $z(k)$ should be such that $E[z(k)v(k)] = 0$

The instrument $z(k)$ should be correlated with $\varphi(k)$

In order to minimize the variance of the estimate, we would like to choose:

$$z(k) = [-y_o(k - 1) \cdots - y_o(k - n_a) u(k - 1) \cdots u(k - n_b)]^\top$$

Summarizing, $z(k)$ should be as much as possible correlated with the noise-free regressor, but not correlated with the residual $v(k)$.

---

**Choice of the instrument: idea**

1. Estimate the model parameters $\hat{\theta}$ through LS (biased estimate!)
2. Perform an open-loop simulation of the estimated model:

$$\hat{y}(k) = [-\hat{y}(k - 1) \cdots - \hat{y}(k - n_a) u(k - 1) \cdots u(k - n_b)]^\top \hat{\theta}$$
Instrumental Variables: how to choose the instruments?

- The instrument $z(k)$ should be such that $\mathbb{E}[z(k)v(k)] = 0$
- The instrument $z(k)$ should be correlated with $\varphi(k)$
- In order to minimize the variance of the estimate, we would like to choose:
  
  $$z(k) = [-y_o(k-1) \cdots - y_o(k-n_a) \ u(k-1) \cdots u(k-n_b)]^\top$$

- Summarizing, $z(k)$ should be as much as possible correlated with the noise-free regressor, but not correlated with the residual $v(k)$.

Choice of the instrument: idea

1. Estimate the model parameters $\hat{\theta}$ through LS (biased estimate!)
2. Perform an open-loop simulation of the estimated model:
   
   $$\hat{y}(k) = [-\hat{y}(k-1) \cdots - \hat{y}(k-n_a) \ u(k-1) \cdots u(k-n_b)]^\top \hat{\theta}$$

3. Construct the instrument $z(k)$ as follows

   $$z(k) = [-\hat{y}(k-1) \cdots - \hat{y}(k-n_a) \ u(k-1) \cdots u(k-n_b)]^\top$$
Instrumental Variables: how to choose the instruments?

- The instrument $z(k)$ should be such that $\mathbb{E}[z(k)\nu(k)] = 0$
- The instrument $z(k)$ should be correlated with $\varphi(k)$
- In order to minimize the variance of the estimate, we would like to choose:
  $$z(k) = [-y_o(k-1) \cdots - y_o(k-n_a) u(k-1) \cdots u(k-n_b)]^\top$$
- Summarizing, $z(k)$ should be as much as possible correlated with the noise-free regressor, but not correlated with the residual $\nu(k)$.

Choice of the instrument: idea

1. Estimate the model parameters $\hat{\theta}$ through LS (biased estimate!)
2. Perform an open-loop simulation of the estimated model:
   $$\hat{y}(k) = [-\hat{y}(k-1) \cdots - \hat{y}(k-n_a) u(k-1) \cdots u(k-n_b)]^\top \hat{\theta}$$
3. Construct the instrument $z(k)$ as follows
   $$z(k) = [-\hat{y}(k-1) \cdots - \hat{y}(k-n_a) u(k-1) \cdots u(k-n_b)]^\top$$
4. Estimate the model parameters $\hat{\theta}$ through IV (consistent estimate)
Instrumental Variables: how to choose the instruments?

- The instrument $z(k)$ should be such that $\mathbb{E}[z(k)v(k)] = 0$
- The instrument $z(k)$ should be correlated with $\varphi(k)$
- In order to minimize the variance of the estimate, we would like to choose:
  \[ z(k) = \left[ -y_o(k-1) \cdots -y_o(k-n_a) \ u(k-1) \cdots u(k-n_b) \right]^T \]
- Summarizing, $z(k)$ should be as much as possible correlated with the noise-free regressor, but not correlated with the residual $v(k)$.

Choice of the instrument: idea

1. Estimate the model parameters $\hat{\theta}$ through LS (biased estimate!)
2. Perform an open-loop simulation of the estimated model:
   \[ \hat{y}(k) = \left[ -\hat{y}(k-1) \cdots -\hat{y}(k-n_a) \ u(k-1) \cdots u(k-n_b) \right]^T \hat{\theta} \]
3. Construct the instrument $z(k)$ as follows
   \[ z(k) = \left[ -\hat{y}(k-1) \cdots -\hat{y}(k-n_a) \ u(k-1) \cdots u(k-n_b) \right]^T \]
4. Estimate the model parameters $\hat{\theta}$ through IV (consistent estimate)
5. Repeat from step 2 until convergence
Prediction Error Methods
\[ y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) \]

\[ \mathbb{E}\left[ e(t)e^\top(s) \right] = \Lambda_e \delta(s - t) \text{ (i.e., } e \text{ is white)} \]

\[ G(0, \theta) = 0, \quad H(0, \theta) = I, \quad H^{-1}(q, \theta) \text{ is stable} \]
Description of Prediction Error Methods

\[ y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) \]

\[ \mathbb{E}\left[e(t)e^\top(s)\right] = \Lambda_e \delta(s - t) \text{ (i.e., } e \text{ is white)} \]

\[ G(0, \theta) = 0, \quad H(0, \theta) = I, \quad H^{-1}(q, \theta) \text{ is stable} \]

Linear predictor

\[ \hat{y}(k|k-1; \theta) = L_y(q^{-1}, \theta)y(k) + L_u(q^{-1}, \theta)u(k) \]

\[ L_y(0, \theta) = 0, \quad L_u(0, \theta) = 0 \]

\[ \hat{y}(k|k - 1; \theta) \text{ only depends on past input/output data} \]
Description of Prediction Error Methods

\[ y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) \]

\[ \mathbb{E}[e(t)e^\top(s)] = \Lambda_e \delta(s - t) \text{ (i.e., } e \text{ is white)} \]

\[ G(0, \theta) = 0, \quad H(0, \theta) = I, \quad H^{-1}(q, \theta) \text{ is stable} \]

**Linear predictor**

\[ \hat{y}(k|k - 1; \theta) = L_y(q^{-1}, \theta)y(k) + L_u(q^{-1}, \theta)u(k) \]

\[ L_y(0, \theta) = 0, \quad L_u(0, \theta) = 0 \]

\[ \hat{y}(k|k - 1; \theta) \text{ only depends on past input/output data} \]

**Prediction error**

\[ \varepsilon(k, \theta) = y(k) - \hat{y}(k|k - 1, \theta) \]
Description of Prediction Error Methods

\[ y(k) = G(q, \theta) u(k) + H(q, \theta) e(k) \]

\[ \mathbb{E} \left[ e(t) e^\top(s) \right] = \Lambda \delta(s - t) \] (i.e., \( e \) is white)

\[ G(0, \theta) = 0, \quad H(0, \theta) = I, \quad H^{-1}(q, \theta) \text{ is stable} \]

**Linear predictor**

\[ \hat{y}(k|k - 1; \theta) = L_y(q^{-1}, \theta) y(k) + L_u(q^{-1}, \theta) u(k) \]

\[ L_y(0, \theta) = 0, \quad L_u(0, \theta) = 0 \]

**Prediction error**

\[ \varepsilon(k, \theta) = y(k) - \hat{y}(k|k - 1, \theta) \]

The estimated parameters \( \hat{\theta} \) should make the prediction errors \( \{\varepsilon(k, \theta)\}_{k=1}^N \) “small”

\[ \hat{y}(k|k - 1; \theta) \text{ only depends on past input/output data} \]
Prediction Error Methods

- Choice of **model structure** (parametrization of $G(q, \theta)$ and $H(q, \theta)$ as a function of $\theta$)
- Choice of **predictor** (define filters of $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$)
- Choice of **criterion** $V_N(\theta)$ (scalar function of the prediction errors $\{\varepsilon(k, \theta)\}^N_{k=1}$ to assess the performance of the predictor)
- Estimate the parameters $\hat{\theta} = \arg \min_{\theta} V_N(\theta)$
SISO model structures

- **AutoRegressive with Exogenous inputs (ARX) models**
  \[ y(k) = \frac{B(q, \theta)}{A(q, \theta)} u(k) + \frac{1}{A(q, \theta)} e(k) \]

- **AutoRegressive-Moving-Average with Exogenous inputs (ARMAX) models**
  \[ y(k) = \frac{B(q, \theta)}{A(q, \theta)} u(k) + \frac{C(q, \theta)}{A(q, \theta)} e(k) \]

- **Output Error (OE) models**
  \[ y(k) = \frac{B(q, \theta)}{A(q, \theta)} u(k) + e(k) \]

- **Box-Jenkins (BJ) models**
  \[ y(k) = \frac{B(q, \theta)}{A(q, \theta)} u(k) + \frac{C(q, \theta)}{D(q, \theta)} e(k) \]
Choice of the predictor filters

**Optimal predictor**

Choose the prediction filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance.
**Choice of the predictor filters**

**Optimal predictor**

Choose the prediction filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance.

**Compute the optimal predictor**

\[
y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) = G(q, \theta)u(k) + (H(q, \theta) - I)e(k) + e(k) = G(q, \theta)u(k) + (H(q, \theta) - I)H^{-1}(q, \theta)(y(k) - G(q, \theta)u(k)) + e(k) =\\
= [(I - H^{-1}(q, \theta))y(k) + H^{-1}(q, \theta)G(q, \theta)u(k)] + e(k) = z(k) + e(k)
\]
Choice of the predictor filters

Optimal predictor

Choose the prediction filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance

Compute the optimal predictor

$$y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) = G(q, \theta)u(k) + (H(q, \theta) - I) e(k) + e(k) =$$
$$= G(q, \theta)u(k) + (H(q, \theta) - I) H^{-1}(q, \theta) (y(k) - G(q, \theta)u(k)) + e(k) =$$
$$= [(I - H^{-1}(q, \theta)) y(k) + H^{-1}(q, \theta)G(q, \theta)u(k)] + e(k) =$$
$$= z(k) + e(k)$$

$z(k)$ and $e(k)$ are uncorrelated
Choice of the predictor filters

**Optimal predictor**

Choose the prediction filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance.

**Compute the optimal predictor**

$$y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) = G(q, \theta)u(k) + (H(q, \theta) - I)e(k) + e(k) =$$

$$= G(q, \theta)u(k) + (H(q, \theta) - I)H^{-1}(q, \theta)(y(k) - G(q, \theta)u(k)) + e(k) =$$

$$= [(I - H^{-1}(q, \theta))y(k) + H^{-1}(q, \theta)G(q, \theta)u(k)] + e(k) =$$

$$= z(k) + e(k)$$

Let $y^*(k)$ be an arbitrary predictor of $y(k)$ based on data up to time $k - 1$.
Choose the predictor filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance

Compute the optimal predictor

\[
y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) = G(q, \theta)u(k) + (H(q, \theta) - I) e(k) + e(k) = \\
= G(q, \theta)u(k) + (H(q, \theta) - I) H^{-1}(q, \theta) (y(k) - G(q, \theta)u(k)) + e(k) = \\
= \left[ (I - H^{-1}(q, \theta)) y(k) + H^{-1}(q, \theta)G(q, \theta)u(k) \right] + e(k) = \\
= z(k) + e(k)
\]

Let $y^*(k)$ be an arbitrary predictor of $y(k)$ based on data up to time $k - 1$

\[
\mathbb{E} \left[ (y(k) - y^*(k)) (y(k) - y^*(k))^\top \right] = \mathbb{E} \left[ (z(k) + e(k) - y^*(k)) (z(k) + e(k) - y^*(k))^\top \right] = \\
= \mathbb{E} \left[ (z(k) - y^*(k)) (z(k) - y^*(k))^\top \right] + \mathbb{E} \left[ e(k)e^\top(k) \right] \succeq \Lambda_e
\]
Choice of the predictor filters

**Optimal predictor**

Choose the prediction filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance.

**Compute the optimal predictor**

\[
y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) = G(q, \theta)u(k) + (H(q, \theta) - I) e(k) + e(k) =
\]
\[
= G(q, \theta)u(k) + (H(q, \theta) - I) H^{-1}(q, \theta) (y(k) - G(q, \theta)u(k)) + e(k) =
\]
\[
= [(1 - H^{-1}(q, \theta)) y(k) + H^{-1}(q, \theta)G(q, \theta)u(k)] + e(k) =~ z(k) + e(k)
\]

Let $y^*(k)$ be an arbitrary predictor of $y(k)$ based on data up to time $k - 1$

\[
\mathbb{E} \left[ (y(k) - y^*(k)) (y(k) - y^*(k))^\top \right] = \mathbb{E} \left[ (z(k) + e(k) - y^*(k)) (z(k) + e(k) - y^*(k))^\top \right] =~
\]
\[
= \mathbb{E} \left[ (z(k) - y^*(k)) (z(k) - y^*(k))^\top \right] + \mathbb{E} \left[ e(k)e^\top(k) \right] \succeq \Lambda_e
\]

$z(k)$ is the optimal predictor and $e(k)$ is the “optimal” prediction error.
Choice of the predictor filters

**Optimal predictor**

Choose the prediction filters $L_y(q^{-1}, \theta)$ and $L_u(q^{-1}, \theta)$ providing the prediction error with smallest variance.

**Compute the optimal predictor**

\[
y(k) = G(q, \theta)u(k) + H(q, \theta)e(k) = G(q, \theta)u(k) + (H(q, \theta) - I)e(k) + e(k) = \\
= G(q, \theta)u(k) + (H(q, \theta) - I)H^{-1}(q, \theta)(y(k) - G(q, \theta)u(k)) + e(k) = \\
= \left[ (I - H^{-1}(q, \theta)) y(k) + H^{-1}(q, \theta)G(q, \theta)u(k) \right] + e(k) = \\
= z(k) + e(k)
\]

Let $y^*(k)$ be an arbitrary predictor of $y(k)$ based on data up to time $k - 1$

\[
\mathbb{E} \left[ (y(k) - y^*(k))(y(k) - y^*(k))^\top \right] = \mathbb{E} \left[ (z(k) + e(k) - y^*(k))(z(k) + e(k) - y^*(k))^\top \right] = \\
= \mathbb{E} \left[ (z(k) - y^*(k))(z(k) - y^*(k))^\top \right] + \mathbb{E} \left[ e(k)e^\top(k) \right] \succeq \Lambda_e
\]

$z(k)$ is the optimal predictor and $e(k)$ is the “optimal” prediction error.

\[
\hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta)G(q^{-1}, \theta)u(k) \\
\varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta)(y(k) - G(q^{-1}, \theta)u(k))
\]
Choice of the loss function $V_N(\theta)$

- Sample covariance matrix

$$R_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k, \theta)\varepsilon^\top(k, \theta)$$
Minimization criterion

Choice of the loss function $V_N(\theta)$

- Sample covariance matrix

\[ R_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \epsilon(k, \theta) \epsilon^\top(k, \theta) \]

- If $y$ is scalar, $R_N(\theta)$ can be taken as the criterion $V_N(\theta)$ to be minimized
Minimization criterion

Choice of the loss function $V_N(\theta)$

- Sample covariance matrix

$$R_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k, \theta)\varepsilon^\top(k, \theta)$$

- If $y$ is scalar, $R_N(\theta)$ can be taken as the criterion $V_N(\theta)$ to be minimized
- In the multivariable case, we can minimize

$$V_N(\theta) = h(R_N(\theta))$$

with $h$ continuous monotonically increasing function defined on the set of positive semidefinite matrices:

$$h(Q + \Delta Q) \geq h(Q) \quad \forall Q, \Delta Q \succeq 0.$$
Minimization criterion

### Choice of the loss function $V_N(\theta)$

- Sample covariance matrix

\[
R_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k, \theta) \varepsilon^\top(k, \theta)
\]

- If $y$ is scalar, $R_N(\theta)$ can be taken as the criterion $V_N(\theta)$ to be minimized

- In the multivariable case, we can minimize

\[
V_N(\theta) = h(R_N(\theta))
\]

with $h$ continuous monotonically increasing function defined on the set of positive semidefinite matrices:

\[
h(Q + \Delta Q) \geq h(Q) \quad \forall Q, \Delta Q \succeq 0.
\]

**Ex:** $V_N(\theta) = h(R_N(\theta)) = tr(R_N(\theta))$
Minimization criterion

**Choice of the loss function $V_N(\theta)$**

- Sample covariance matrix

\[
R_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k, \theta)\varepsilon^\top(k, \theta)
\]

- If $y$ is scalar, $R_N(\theta)$ can be taken as the criterion $V_N(\theta)$ to be minimized

- In the multivariable case, we can minimize

\[
V_N(\theta) = h(R_N(\theta))
\]

with $h$ continuous monotonically increasing function defined on the set of positive semidefinite matrices:

\[
h(Q + \Delta Q) \geq h(Q) \quad \forall Q, \Delta Q \succeq 0.
\]

**Ex:** $V_N(\theta) = h(R_N(\theta)) = tr\left(R_N(\theta)\right)$

**Final estimate**

\[
\hat{\theta}_{PEM} = \arg\min_{\theta} V_N(\theta)
\]
What happens when $N \to \infty$?

- $\lim_{N \to \infty} R_N(\theta) = \mathbb{E} \left[ \varepsilon(k, \theta) \varepsilon^\top(k, \theta) \right] = R_\infty(\theta)$
- $\lim_{N \to \infty} h(R_N(\theta)) = h(R_\infty(\theta)) = V_\infty(\theta)$
- Convergence is uniform on a compact set $\Theta$, i.e.,
  \[ \sup_{\theta \in \Theta} |V_N(\theta) - V_\infty(\theta)| \to 0 \]
- $\lim_{N \to \infty} \hat{\theta}_{PEM} = \theta^* = \arg\min_{\theta} V_\infty(\theta)$
Is $\hat{\theta}_{\text{PEM}}$ a consistent estimate of $\theta_0$?

- Let $\theta_0$ be the true system parameters:

$$y(k) = G(q, \theta_0)u(k) + H(q, \theta_0)e(k), \quad \mathbb{E} \left[ e(t)e^\top(s) \right] = \Lambda_e \delta(s - t),$$

$$G(0, \theta_0) = 0, \quad H(0, \theta_0) = I, \quad H^{-1}(q, \theta_0) \text{ stable}$$

- Thus:

$$\varepsilon(k, \theta) = H^{-1}(q, \theta) (G(q, \theta_0)u(k) + H(q, \theta_0)e(k) - G(q, \theta)u(k)) =$$

$$= H^{-1}(q, \theta) (G(q, \theta_0) - G(q, \theta)) u(k) + H^{-1}(q, \theta)H(q, \theta_0)e(k) =$$

$$= e(k) + \text{ terms independent of } e(k)$$

- Thus: $R_\infty(\theta) = \mathbb{E} \left[ \varepsilon(k, \theta)\varepsilon^\top(k, \theta) \right] \geq \mathbb{E} \left[ e(k)e^\top(k) \right] = \Lambda_e$

- $\theta_0$ is a minimizer of $h(R_\infty(\theta)) = V_\infty$.

- If $u(k)$ and $e(k)$ are not correlated, only the “true” parameters $\theta_0$ minimize $h(R_\infty(\theta)) = V_\infty$

- Thus $\lim_{N \to \infty} \hat{\theta}_{\text{PEM}} = \theta_0$. 
\[ \hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k) \]
\[ \varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k)) \]

**Predictor for ARX models**

\[ y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{1}{A(q^{-1}, \theta)} e(k), \quad e \text{ white} \]
\[
\hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k)
\]

\[
\varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k))
\]

**Predictor for ARX models**

\[
y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{1}{A(q^{-1}, \theta)} e(k), \quad e \text{ white}
\]

- **Optimal predictor:**

\[
\hat{y}(k|k-1, \theta) = (I - A(q^{-1}, \theta)) y(k) + B(q^{-1}, \theta) u(k)
\]

\[
= -a_1 y(k-1) - \cdots - a_{n_a} y(k-n_a) + b_1 u(k-1) + \cdots + b_{n_b} u(k-n_b) = 
\]

\[
= \varphi^T(k) \theta
\]
\[ \hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k) \]
\[ \varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k)) \]

Predictor for ARX models

\[
y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{1}{A(q^{-1}, \theta)} e(k), \quad e \text{ white}
\]

**Optimal predictor:**

\[
\hat{y}(k|k-1, \theta) = (I - A(q^{-1}, \theta)) y(k) + B(q^{-1}, \theta) u(k)
\]
\[= -a_1 y(k-1) - \cdots - a_{n_a} y(k-n_a) + b_1 u(k-1) + \cdots + b_{n_b} u(k-n_b) = \]
\[= \varphi^\top(k) \theta \quad \text{Linear regression representation} \]
\[ \hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k) \]
\[ \varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k)) \]

**Predictor for ARX models**

\[ y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{1}{A(q^{-1}, \theta)} e(k), \quad e \text{ white} \]

- **Optimal predictor:**

  \[
  \hat{y}(k|k-1, \theta) = (I - A(q^{-1}, \theta)) y(k) + B(q^{-1}, \theta) u(k)
  \]
  \[
  = -a_1 y(k - 1) - \cdots - a_{n_a} y(k - n_a) + b_1 u(k - 1) + \cdots + b_{n_b} u(k - n_b) = \\
  = \phi^\top(k) \theta
  \]

- **PEM estimate:**

  \[
  \hat{\theta}_{PEM} = \min_{\theta} \sum_{k=1}^{N} (y(k) - \hat{y}(k|k-1, \theta))^2
  \]
\[
\hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta))y(k) + H^{-1}(q^{-1}, \theta)G(q^{-1}, \theta)u(k)
\]
\[
\varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta)(y(k) - G(q^{-1}, \theta)u(k))
\]

** Predictor for ARX models **

\[
y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{1}{A(q^{-1}, \theta)} e(k), \quad e \text{ white}
\]

- **Optimal predictor:**

\[
\hat{y}(k|k-1, \theta) = (I - A(q^{-1}, \theta))y(k) + B(q^{-1}, \theta)u(k)
\]

\[
= -a_1 y(k-1) - \cdots - a_{n_a} y(k-n_a) + b_1 u(k-1) + \cdots + b_{n_b} u(k-n_b) = \varphi^\top(k) \theta
\]

- **PEM estimate:**

\[
\hat{\theta}_{\text{PEM}} = \min_{\theta} \sum_{k=1}^{N} (y(k) - \hat{y}(k|k-1, \theta))^2 = \min_{\theta} \sum_{k=1}^{N} \left(y(k) - \varphi^\top(k) \theta\right)^2 = \left(\Phi^\top \Phi\right)^{-1} \Phi^\top Y
\]
\[
\hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k)
\]
\[
\varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k))
\]

**Predictor for ARX models**

\[
y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{1}{A(q^{-1}, \theta)} e(k), \quad e \text{ white}
\]

- **Optimal predictor:**
  \[
  \hat{y}(k|k-1, \theta) = (I - A(q^{-1}, \theta)) y(k) + B(q^{-1}, \theta) u(k)
  \]
  \[
  = -a_1 y(k-1) - \cdots - a_{n_a} y(k-n_a) + b_1 u(k-1) + \cdots + b_{n_b} u(k-n_b) = \varphi^\top(k) \theta
  \]
  Linear regression representation

- **PEM estimate:**
  \[
  \hat{\theta}_{\text{PEM}} = \min_{\theta} \sum_{k=1}^{N} (y(k) - \hat{y}(k|k-1, \theta))^2 = \min_{\theta} \sum_{k=1}^{N} \left( y(k) - \varphi^\top(k) \theta \right)^2 = \left( \Phi^\top \Phi \right)^{-1} \Phi^\top Y
  \]
  LS method and PEM coincide!
\[ \hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k) \]

\[ \epsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k)) \]

Predictor for OE models

\[ y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + e(k), \quad e \text{ white} \]
\[
\hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k)
\]
\[
\epsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k))
\]

**Predictor for OE models**

\[
y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + e(k), \quad e \text{ white}
\]

- Optimal predictor:

\[
\hat{y}(k|k-1, \theta) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) =
\]
\[
= -a_1 \hat{y}(k-1|k-2) - \cdots - a_{n_a} \hat{y}(k-n_a|k-n_a-1) +
\]
\[
+ b_1 u(k-1) + \cdots + b_{n_b} u(k-n_b) =
\]
\[
= \hat{\phi}^\top(k) \theta
\]
\[
\hat{y}(k|k-1, \theta) = (I - H^{-1}(q^{-1}, \theta)) y(k) + H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) u(k) \\
\varepsilon(k, \theta) = e(k) = H^{-1}(q^{-1}, \theta) (y(k) - G(q^{-1}, \theta) u(k))
\]
PEM: numerical optimization

- PEM estimate:
  \[ \hat{\theta}_{\text{PEM}} = \arg \min_{\theta} V_N(\theta) = \arg \min_{\theta} h \left( \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k, \theta) \varepsilon^\top(k, \theta) \right) \]

- The solution cannot be always computed analytically

- Numerical iterative algorithms for non-convex optimization should be used:
  
  i. initialize with an initial estimate \( \hat{\theta}^{(0)} \)
  
  ii. update: \( \hat{\theta}^{(i+1)} = f(\hat{\theta}^{(i)}) \) (the estimate is iteratively refined)
  
  iii. we would like that the estimate converges to the optimum \( \hat{\theta}_{\text{PEM}} \):

\[ \hat{\theta}^{(0)} \rightarrow \hat{\theta}^{(1)} \rightarrow \hat{\theta}^{(2)} \rightarrow \cdots \rightarrow \hat{\theta}_{\text{PEM}} \]
Gradient method

- choose an initial condition $\hat{\theta}^{(0)}$;
- iterate
  
  (i) line search: choose a positive step size $t > 0$
  
  (ii) update: $\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - t \left. \frac{\partial V_N(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}^{(i)}}$

- until stopping criterion is satisfied (typically: $\left\| \frac{\partial V_N(\theta)}{\partial \theta} \right\| \leq \epsilon$)

- in case of scalar output, $V_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon^2(k, \theta)$ and $\frac{\partial V_N(\theta)}{\partial \theta} = \frac{2}{N} \sum_{k=1}^{N} \varepsilon(k, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial \theta}$

- it converges (slowly) to the global optimum if $V_N(\theta)$ is convex

- in case of non-convex $V_N(\theta)$, convergence to the global minimum is not guaranteed

Exact line search

$$t = \arg\min_{t>0} V_N \left( \hat{\theta}^{(i)} + t\Delta \theta \right), \text{ with } \Delta \theta = - \left. \frac{\partial V_N(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}^{(i)}}$$
choose an initial condition \( \hat{\theta}^{(0)} \);

iterate

(i) line search: choose a positive step size \( t > 0 \)
(ii) update: \( \hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - t \left( \nabla^2 V_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla V_N \left( \hat{\theta}^{(i)} \right) \)

until stopping criterion is satisfied. Typically: \( \left| \nabla V_N(\hat{\theta}^{(i)})^\top \left( \nabla^2 V_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla V_N \left( \hat{\theta}^{(i)} \right) \right| \leq \epsilon \)

in case of scalar output, \( V_N(\theta) = \frac{1}{N} \sum_{k=1}^{N} \epsilon^2(k, \theta) \) and

\[
\nabla V_N(\theta) = \frac{2}{N} \sum_{k=1}^{N} \epsilon(k, \theta) \frac{\partial \epsilon(k, \theta)}{\partial \theta}, \quad \nabla^2 V_N(\theta) = \frac{2}{N} \sum_{k=1}^{N} \frac{\partial \epsilon(k, \theta)}{\partial \theta} \frac{\partial^\top \epsilon(k, \theta)}{\partial \theta} + \frac{2}{N} \sum_{k=1}^{N} \frac{\partial^2 \epsilon(k, \theta)}{\partial \theta^2} \epsilon(k, \theta)
\]
\[ \hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + t \Delta \theta, \text{ with } \Delta \theta = - \left( \nabla^2 V_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla V_N(\hat{\theta}^{(i)}) \]

**Gauss-Newton method: interpretation**

\[ \hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + \Delta \theta \] minimizes the second order approximation:

\[ \hat{V}_N(\theta^{(i)} + \Delta \theta) = V_N(\theta^{(i)}) + \nabla V_N(\theta^{(i)})^\top \Delta \theta + \frac{1}{2} \Delta \theta^\top \nabla^2 V_N(\theta^{(i)}) \Delta \theta \]

The minimum of the quadratic function above is achieved at

\[ \Delta \theta = - \left( \nabla^2 V_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla V_N(\hat{\theta}^{(i)}) \]
\[
\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + t \Delta \theta, \quad \text{with} \quad \Delta \theta = - \left( \nabla^2 \mathcal{V}_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla \mathcal{V}_N(\hat{\theta}^{(i)})
\]

Gauss-Newton method: interpretation

\[
\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + \Delta \theta \quad \text{minimizes the second order approximation:}
\]

\[
\hat{\mathcal{V}}_N(\theta^{(i)} + \Delta \theta) = \mathcal{V}_N(\theta^{(i)}) + \nabla \mathcal{V}_N(\theta^{(i)})^\top \Delta \theta + \frac{1}{2} \Delta \theta^\top \nabla^2 \mathcal{V}_N(\theta^{(i)}) \Delta \theta
\]

The minimum of the quadratic function above is achieved at

\[
\Delta \theta = - \left( \nabla^2 \mathcal{V}_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla \mathcal{V}_N(\hat{\theta}^{(i)})
\]

Careful: if the Hessian is not positive definite, we move to the “wrong” direction.
\[
\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + t \Delta \theta, \quad \text{with} \quad \Delta \theta = - \left( \nabla^2 V_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla V_N(\hat{\theta}^{(i)})
\]

**Gauss-Newton method: interpretation**

\( \hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + \Delta \theta \) minimizes the second order approximation:

\[
\hat{V}_N(\theta^{(i)} + \Delta \theta) = V_N(\theta^{(i)}) + \nabla V_N(\theta^{(i)})^\top \Delta \theta + \frac{1}{2} \Delta \theta^\top \nabla^2 V_N(\theta^{(i)}) \Delta \theta
\]

The minimum of the quadratic function above is achieved at:

\[
\Delta \theta = - \left( \nabla^2 V_N(\hat{\theta}^{(i)}) \right)^{-1} \nabla V_N(\hat{\theta}^{(i)})
\]

Careful: if the Hessian is not positive definite, we move to the “wrong” direction.

Hessian approximation:

\[
\nabla^2 V_N(\theta) = \frac{2}{N} \sum_{k=1}^{N} \frac{\partial \varepsilon(k, \theta)}{\partial \theta} \frac{\partial \varepsilon^\top(k, \theta)}{\partial \theta} + \frac{2}{N} \sum_{k=1}^{N} \frac{\partial^2 \varepsilon(k, \theta)}{\partial \theta^2} \varepsilon(k, \theta) \approx \frac{2}{N} \sum_{k=1}^{N} \frac{\partial \varepsilon(k, \theta)}{\partial \theta} \frac{\partial \varepsilon^\top(k, \theta)}{\partial \theta} + \frac{\delta I}{\Delta I} \geq 0
\]

regularization
Example: evaluation of the gradient for ARMAX models

- ARMAX model: 
  \[ y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{C(q^{-1}, \theta)}{A(q^{-1}, \theta)} e(k) \]

- Prediction error \( \varepsilon(k, \theta) \):
  \[ C(q^{-1}, \theta) \varepsilon(k, \theta) = A(q^{-1}, \theta) y(k) - B(q^{-1}, \theta) u(k) \]
Example: evaluation of the gradient for ARMAX models

- ARMAX model:  
  \[ y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{C(q^{-1}, \theta)}{A(q^{-1}, \theta)} e(k) \]

- Prediction error: \[ \varepsilon(k, \theta) : C(q^{-1}, \theta)\varepsilon(k, \theta) = A(q^{-1}, \theta)y(k) - B(q^{-1}, \theta)u(k) \]

- Compute derivatives of both left and right hand of the above equation:

  \[
  C(q^{-1}, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial a_i} = y(k - i) \Rightarrow \frac{\partial \varepsilon(k, \theta)}{\partial a_i} = \frac{1}{C(q^{-1}, \theta)} y(k - i)
  \]

  \[
  C(q^{-1}, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial b_i} = -u(k - i) \Rightarrow \frac{\partial \varepsilon(k, \theta)}{\partial b_i} = -\frac{1}{C(q^{-1}, \theta)} u(k - i)
  \]

  \[
  \varepsilon(k - i, \theta) + C(q^{-1}, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial c_i} = 0 \Rightarrow \frac{\partial \varepsilon(k, \theta)}{\partial c_i} = -\frac{1}{C(q^{-1}, \theta)} \varepsilon(k - i, \theta)
  \]
Example: evaluation of the gradient for ARMAX models

- **ARMAX model:**
  \[
  y(k) = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u(k) + \frac{C(q^{-1}, \theta)}{A(q^{-1}, \theta)} e(k)
  \]

- **Prediction error:**
  \[
  \varepsilon(k, \theta) : C(q^{-1}, \theta)\varepsilon(k, \theta) = A(q^{-1}, \theta)y(k) - B(q^{-1}, \theta)u(k)
  \]

- Compute derivatives of both left and right hand of the above equation:

  \[
  C(q^{-1}, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial a_i} = y(k - i) \Rightarrow \frac{\partial \varepsilon(k, \theta)}{\partial a_i} = \frac{1}{C(q^{-1}, \theta)} y(k - i)
  \]

  \[
  C(q^{-1}, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial b_i} = -u(k - i) \Rightarrow \frac{\partial \varepsilon(k, \theta)}{\partial b_i} = -\frac{1}{C(q^{-1}, \theta)} u(k - i)
  \]

  \[
  \varepsilon(k - i, \theta) + C(q^{-1}, \theta) \frac{\partial \varepsilon(k, \theta)}{\partial c_i} = 0 \Rightarrow \frac{\partial \varepsilon(k, \theta)}{\partial c_i} = -\frac{1}{C(q^{-1}, \theta)} \varepsilon(k - i, \theta)
  \]

- Thus:

  \[
  \frac{\partial \varepsilon(k, \theta)}{\partial \theta} = [y_F(k - 1) \cdots y_F(k - n_a) - u_F(k - 1) \cdots - u_F(k - n_b) - \varepsilon_F(k - 1) \cdots - \varepsilon_F(k - n_c)]^T
  \]

  with:

  \[
  y_F(k) = \frac{1}{C(q^{-1}, \theta)} y(k), \quad u_F(k) = \frac{1}{C(q^{-1}, \theta)} u(k), \quad \varepsilon_F(k) = \frac{1}{C(q^{-1}, \theta)} \varepsilon(k)
  \]